https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

Accelerating greenhouse gas retrievals with neural network-based forward models

Fiona Lippert[1], Andrew Gerald Barr[1], Marcos Herreras-Giralda[2], Masahiro Momoi[2], Fernando Rejano[2], Sha Lu[1], Otto Hasekamp[1], Oleg Dubovik[3], Edward Malina[4], and Jochen Landgraf[1]

[1]Space Research Organisation Netherlands (SRON), Leiden, Netherlands
[2]GRASP SAS, Lille, France
[3]Univ. Lille, CNRS, UMR 8518 - LOA - Laboratoire d'Optique Atmosphérique, Lille, France
[4]ESA/ESRIN, Frascati, Italy

Correspondence: Fiona Lippert (f.lippert@sronl.nl)

Abstract. Greenhouse gas (GHG) retrievals rely on repeated evaluations of computationally expensive physics-based forward models, which limit the feasibility of near-real-time retrievals and timely detection of emission hotspots. A promising alternative is to replace these forward models with fast machine learning emulators trained to approximate their input-output mapping, while retaining the overall retrieval algorithm. Here, we assess the feasibility of this approach in the context of the Sentinel-5 mission, systematically comparing two different emulation strategies: an end-to-end approach, which directly approximates the full forward model with neural networks, and a hybrid approach, which combines fast non-scattering simulations with a neural network-based correction for atmospheric scattering effects. We comprehensively validate each emulator in the full retrieval chain, evaluating their impact on the accuracy of retrieved XCO_2 and XCH_4. Our results show that a hybrid approach is needed to meet the stringent accuracy requirements on XCH_4 and XCO_2. While the end-to-end emulator achieves large speed-ups exceeding a factor of 300, it introduces considerable errors of 7.22 ppb for XCH_4 and 4.25 ppm for XCO_2 compared to full-physics retrievals, and fails to generalize to high-emission scenarios beyond the training range. In contrast, the hybrid approach can effectively leverage the information provided by the non-scattering approximation, reducing emulator-induced retrieval errors to less than 1.5 ppb for XCH_4 and 0.5 ppm for XCO_2, while still being an order of magnitude faster than full-physics retrievals and maintaining robust performance for high-emission scenarios. Together, these results pave the way for operational deployment of neural network-based forward models in GHG retrievals from Sentinel-5, and more broadly demonstrate the potential of hybrid machine learning emulators to facilitate timely and accurate processing of the rapidly growing data volumes from modern satellite missions.

1 Introduction

Accurate measurements of greenhouse gas concentrations are essential for understanding the Earth's carbon cycle, improving climate projections, and assessing the effectiveness of emission reduction policies (IPCC, 2021; Friedlingstein et al., 2023). While surface-based monitoring networks provide continuous high-precision measurements (Wunch et al., 2011; Heiskanen et al., 2022), their sparse spatial coverage limits the ability to capture regional variability and attribute emissions to specific

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

sources. Satellite remote sensing has become a powerful complementary tool in this regard. Missions such as GOSAT and GOSAT-2, OCO-2/3, Sentinel-5P/TROPOMI, Sentinel-5, GOSAT-GW, and the upcoming CO2M mission provide near-global coverage at increasingly high spatial resolutions, thereby enabling detailed characterization of the spatiotemporal variability of greenhouse gas concentrations (Eldering et al., 2017; Hu et al., 2018) and automated detection of emission hotspots (Schuit et al., 2023).

These satellite missions rely on high-resolution spectrometers that measure the sunlight backscattered by the Earth and its atmosphere across broad spectral ranges in the near-infrared (NIR) and shortwave infrared (SWIR), capturing characteristic absorption features of carbon dioxide (CO_2), methane (CH_4), and other trace gases. The measured radiance spectra are translated into estimates of the underlying gas concentrations using atmospheric retrieval algorithms. This typically involves iterative optimization approaches, in which a forward model is repeatedly evaluated against the observations while adjusting model parameters that characterize atmospheric composition and surface properties (Rodgers, 2000).

Importantly, to facilitate reliable estimation of greenhouse gas sources and sinks from satellite measurements, errors in retrieved column-averaged dry-air mole fractions of CH_4 (XCH_4) and CO_2 (XCO_2) must remain within approximately 1 % to resolve relatively small deviations from the atmospheric background (Meirink et al., 2006; Chevallier et al., 2007). To meet these stringent requirements, it is essential that the forward model accurately simulates radiative transfer through the atmosphere, accounting for absorption and scattering by gases, aerosols and clouds (Butz et al., 2009). This involves computationally expensive line-by-line radiative transfer calculations across thousands of wavelengths to resolve relevant absorption features (Clough et al., 2005; Gordon et al., 2022), combined with complex multiple-scattering simulations to capture modifications to the light path (Hansen and Travis, 1974). Repeated evaluations of the forward model and its derivatives during each retrieval iteration and for every spatial pixel further exacerbate this computational complexity. As a result, forward models present a major computational bottleneck for greenhouse gas retrievals, hampering the implementation of near-real-time pipelines and the processing of rapidly growing data volumes from modern satellite missions.

Reducing the computational cost of full-physics radiative transfer calculations has been an active area of research for decades and a broad range of methods has been developed for this purpose. Correlated k-distribution methods (Lacis and Oinas, 1991; Fu and Liou, 1992), principal component analysis (PCA) based compression schemes (Liu et al., 2006; Efremenko et al., 2014), and optimal spectral sampling (Moncet et al., 2008) reduce the number of spectral calculations by exploiting the inherent redundancy in high-dimensional radiance spectra. The linear-k method (Hasekamp and Butz, 2008) further accelerates computations by decoupling gas absorption from multiple-scattering calculations, significantly reducing the number of expensive scattering simulations required. While these methods have enabled more efficient retrievals, they still retain the full scattering physics and fundamental trade-offs persist between computational speed and accuracy.

Alternatively, scattering can be neglected altogether. This reduces the radiative transfer problem to much cheaper non-scattering calculations, but can introduce large systematic errors in the retrieved gas concentrations, particularly in the presence of aerosols and clouds (Butz et al., 2009). The proxy method (Frankenberg et al., 2005) mitigates this problem by retrieving a target gas relative to a simultaneously measured reference gas, assuming that both are subject to similar light path modifications. While effective, this approach is inherently limited to instruments measuring a suitable reference gas in a nearby spectral

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

region comparable absorption strength, which restricts its applicability to a small number of satellite missions and trace gases. Moreover, the proxy method requires that the reference gas concentration is stable and known a priori, an assumption that breaks down if both gases are co-emitted by the same or nearby sources and precludes independent retrievals of both gases.

A promising alternative strategy to accelerate atmospheric retrievals is to replace computationally expensive physics-based forward models with fast machine learning emulators, which are trained on a large representative dataset of full-physics simulations. In contrast to direct inversion approaches (David et al., 2021), which replace the retrieval itself with a machine learning model, such emulator-based retrievals preserve the full numerical inversion framework and naturally provide important diagnostics such as averaging kernels and propagated error estimates. They can therefore be readily integrated into existing data processing and analysis workflows.

Previous work on machine learning forward model emulators has focused primarily on aerosol and surface retrieval applications, where several studies have demonstrated that both neural network-based and Gaussian process-based emulators can approximate physics-based forward models with sufficient accuracy for operational use, while achieving speed-ups of several orders of magnitude. These studies broadly follow two different emulation strategies: end-to-end approaches, which directly approximate the full forward model from input-output examples alone (Nanda et al., 2019; Gao et al., 2021; Bue et al., 2019; Gómez-Dans et al., 2016), and hybrid approaches, which integrate machine learning models with selected physical components to leverage prior knowledge and reduce the complexity of the learning problem (Brodrick et al., 2021; Fan et al., 2019).

Yet, developing such approaches for operational greenhouse gas retrievals poses a more fundamental challenge. The forward model mapping is considerably more complex, with information encoded in subtle high-resolution absorption features. And the tight accuracy requirements on the retrieved gas concentrations leave little margin for emulator-induced spectral errors. As a result, the use of machine learning emulators in greenhouse gas retrievals remains relatively unexplored. Only a few studies have investigated end-to-end and hybrid approaches for this purpose (Mauceri et al., 2022; Brence et al., 2023; Lamminpää et al., 2025), and most of these have focused on spectral accuracy rather than comprehensive evaluation within a retrieval framework. Consequently, the impact of emulator-induced spectral errors on retrieval performance, as well as the relative benefits of hybrid versus end-to-end approaches, remain insufficiently understood.

In this study, we therefore develop end-to-end and hybrid neural network-based emulators for high-resolution spectrometer forward models and systematically evaluate their performance in terms of computational efficiency, spectral accuracy, and retrieval quality. To this end, we use synthetic Sentinel-5 measurements with known atmospheric states and a consistent forward model for both training and evaluation, enabling us to isolate the impact of emulator-induced forward model errors on retrieved XCH_4 and XCO_2. Our hybrid approach is inspired by the successive orders of scattering theory (Lenoble et al., 1985), which decomposes the total radiance into contributions from increasing orders of scattering, each depending on the previous as its source function. We exploit this natural decomposition by using non-scattering simulations as a fast physics-based backbone and training neural networks to correct for the discarded scattering contributions. This results in a substantially simpler learning problem as the backbone already captures the dominant spectral structure, including molecular absorption features. Moreover, being grounded in well-established radiative transfer theory, the backbone remains informative beyond the training distribution. The resulting hybrid emulators are therefore expected to be more data efficient and extrapolate more reliably compared to

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

purely data-driven approaches. To test this, we investigate the effects of training data size on spectral accuracy and assess how well emulator-based retrievals generalize to high-emission scenarios with gas concentrations exceeding the training range. Together, our analyses shed light on the practical feasibility and limitations of neural network-based emulators for greenhouse gas retrievals and provide guidance on the design of fast, accurate, and robust forward model surrogates.

2 Methodology

We aim to accelerate greenhouse gas retrievals by replacing traditional physics-based forward models with fast machine learning emulators, while keeping the overall retrieval algorithm intact. We develop and evaluate this approach in the context of XCH_4 and XCO_2 retrievals from synthetic Sentinel-5 measurements. The following section presents this approach in more detail, covering instrument specifications (Section 2.1), the associated retrieval algorithm (Section 2.2), the full-physics forward model used as a reference and for generating training data (Section 2.3), the non-scattering approximation forming the backbone of our hybrid approach (Section 2.4), the design and training of our machine learning emulators (Section 2.5-2.7), and the experimental setup used for evaluation (Section 2.8).

2.1 The Sentinel-5 mission and instrument specifications

The Sentinel-5 (S5) mission, launched in August 2025 aboard the MetOp-SG-A1 satellite, is the successor to the Sentinel-5 Precursor (Sentinel-5P) mission and forms part of the European Copernicus observation program. Its core instrument, the Ultraviolet, Visible, Near-infrared, and Shortwave Infrared (UVNS) spectrometer, covers seven spectral bands from 270 nm to 2385 nm and provides near-daily global coverage of atmospheric composition at a spatial sampling of approximately $7.1 \times 7.5\ \mathrm{km}^2$ at nadir (Ingmann et al., 2012; Sierk et al., 2018; Irizar et al., 2019). In this study, we focus on the NIR, SWIR1, and SWIR3 bands of the S5/UVNS spectrometer, which cover characteristic absorption features of O_2, CO_2, CH_4, CO, and H_2O (see Table 1). The spectral resolution of each band is defined by the full width at half maximum (FWHM) of the instrument spectral response function (ISRF), and the associated measurement precision is characterized by the signal-to-noise ratio:

$$\mathrm{SNR} = \frac{aI}{\sqrt{aI + b}}, \tag{1}$$

where a and b are spectrometer-specific noise parameters. The spectral properties and noise parameters for each S5/UVNS band considered in this study are based on the Sentinel-5 instrument specifications (Langen et al., 2011) and are summarized in Table 1.

2.2 Retrieval algorithm

We retrieve XCH_4 and XCO_2 from synthetic S5/UVNS measurements using the RemoTAP retrieval framework developed at SRON, which implements the RemoTeC algorithm (Butz et al., 2009, 2010; Schepers et al., 2014) for spectrometer-only gas retrievals. RemoTeC has been successfully used for XCH_4 and XCO_2 retrievals from GOSAT (Butz et al., 2011; Guerlet et al.,

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

Table 1. Instrument specifications used to generate synthetic S5/UVNS spectrometer measurements.

band	wavelength range (nm)	resolution (nm)	sampling ratio	number of channels	noise parameter a (ph^{-1} cm^2 s nm sr)	noise parameter b (1)	molecular absorption
NIR	$745-773$	0.4	3	211	3.67×10^{-8}	$35,294$	O_2
SWIR1	$1590-1675$	0.25	2.5	851	6.73×10^{-8}	$37,133$	CH_4, CO_2, H_2O
SWIR3	$2305-2385$	0.25	2.5	801	1.20×10^{-7}	$44,880$	CH_4, CO, H_2O

2013) and TROPOMI (Hu et al., 2018; Lorente et al., 2021), and is the algorithm implemented for the operational Sentinel-5 CH_4 product (Landgraf et al., 2025).

Given a vector of spectrometer measurements $\mathbf{y}$, the retrieval algorithm infers a state vector $\mathbf{x}$ of atmospheric and surface properties via iterative optimization. The state vector is linked to the measurements through the forward model $F(\mathbf{x},\mathbf{b})$, which simulates the true atmospheric and measurement processes as a function of $\mathbf{x}$ and ancillary (fixed) parameters $\mathbf{b}$ up to forward model error $\mathbf{e}_F$ and measurement noise $\mathbf{e}_y$:

$$\mathbf{y} = F(\mathbf{x},\mathbf{b}) + \mathbf{e}_F + \mathbf{e}_y, \tag{2}$$

The inverse problem of estimating $\mathbf{x}$ from $\mathbf{y}$ is solved with the Tikhonov regularization method (Phillips, 1962; Tikhonov, 1963), which finds the optimal state vector by minimizing the following cost function:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} ||\mathbf{S}_y^{-1/2}\left(F(\mathbf{x},\mathbf{b})-\mathbf{y}\right)||^2 + \gamma||\mathbf{W}\left(\mathbf{x}-\mathbf{x}_a\right)||^2, \tag{3}$$

where $\mathbf{x}_a$ is the a priori state vector, and $\mathbf{S}_y$ is the measurement error covariance matrix. The regularization parameter γ defines the strength of the side constraint and the diagonal weighting matrix $\mathbf{W}$ determines the relative contribution of each parameter. Since the forward model F is generally non-linear, the optimization is performed iteratively, where at each iteration i the forward model is linearized around the current estimate $\mathbf{x}_i$ using the Jacobian of the forward model at $\mathbf{x}_i$. The next estimate is determined using a Gauss-Newton scheme with reduced step size to avoid diverging retrievals (Hasekamp et al., 2011).

To meet the stringent accuracy requirements for XCH_4 and XCO_2, modifications of the effective light path due to particle scattering by aerosols and clouds must be accounted for in the retrieval algorithm (Butz et al., 2009). Therefore, RemoTeC implements a "full-physics" retrieval approach that simultaneously retrieves gas concentrations and scattering properties of the atmosphere.

2.3 Full-physics forward model

Full-physics retrievals require a forward model capable of accurately simulating radiative transfer in scattering atmospheres. Here, we adopt the full-physics forward model formulation described by Butz et al. (2009, 2010), which relies on few effective aerosol parameters and an efficient multiple-scattering approximation to balance the accuracy and computational cost of evaluating F and its Jacobian at each retrieval iteration.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

At the core of this forward model is the LINTRAN radiative transfer model (RTM) (Landgraf et al., 2002; Hasekamp and Landgraf, 2005; Schepers et al., 2014). It calculates the solar radiation reaching the satellite instrument after reflection by the Earth surface, accounting for absorption and scattering by gas molecules as well as particles in the atmosphere. Here, we use the scalar version of LINTRAN, calculating top-of-the-atmosphere radiances without considering polarization. Direct transmission and single-scattering contributions are calculated directly on a very fine wavelength grid to resolve individual absorption lines, while higher-order scattering contributions are approximated with the linear-k method (Hasekamp and Butz, 2008) to reduce the number of costly multiple-scattering calculations. Jacobians are computed in a numerically efficient manner using forward-adjoint perturbation theory.

The optical properties used to solve the radiative transfer equation are determined based on the thermodynamic state and composition of a vertically discretized model atmosphere. In this study, the atmosphere is divided into 20 homogeneous layers that are equidistant in pressure, with the lowest layer corresponding to the surface pressure. We assume a cloud-free atmosphere consisting only of trace gases and aerosols, where each layer is characterized by its absorption optical thickness, scattering optical thickness, and scattering phase function.

All gases are modeled with a fixed reference vertical profile that is scaled by a single multiplicative factor to adjust the total column concentration during the retrieval. This profile scaling approach has been shown to be fast, robust, and sufficient to reach the required accuracy required for greenhouse gas retrievals (Borsdorff et al., 2014).

Aerosols are represented using a simple parametric model, assuming a power-law size distribution

$$n_{aer}(r) = \begin{cases} C & \text{for } r \le r_1 \\ C\left(\frac{r}{r_1}\right)^{-p} & \text{for } r_1 < r \le r_2 \\ 0 & \text{for } r > r_2 \end{cases} \tag{4}$$

as a function of particle radius r, with fixed cut-offs $r_1 = 0.1$ µm and $r_2 = 10.0$ µm, and normalization constant C. All particles are assumed to be spherical, and the real and imaginary component of the refractive index are fixed to $m_r = 1.4$ and $m_i = -0.003$ across all wavelengths. The vertical distribution of aerosol particles is modeled as a Gaussian with mean z_{aer} and fixed standard deviation $w_{aer} = 2$ km. The aerosol size parameter p, the aerosol central height z_{aer}, and the aerosol column number N_{aer} are the only free parameters of this aerosol model (Butz et al., 2009).

The surface reflectivity is described by a Lambertian albedo. For simplicity, we only consider land surfaces and assign a single constant albedo to each spectral band of the spectrometer, neglecting any fine-scale variation within the band.

Finally, the high-resolution radiances $I(\lambda)$ computed by the RTM are convolved with the ISRF (see Section 2.1) to obtain synthetic radiance measurements for each instrument channel k:

$$I_k = \int_{\lambda_{\min}}^{\lambda_{\max}} f_{\text{instr}}(\lambda_k, \lambda_k - \lambda) I(\lambda) d\lambda. \tag{5}$$

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

2.4 Non-scattering forward model

A computationally attractive simplification of the full-physics forward model is to neglect scattering altogether, reducing the radiative transfer problem to a simple application of the Beer-Lambert law. Under this approximation, top-of-atmosphere radiances are computed as

$$\tilde{I}(\lambda) = A(\lambda) \cdot F_0(\lambda) \cdot \frac{\cos(\theta_0)}{\pi} \cdot \exp(-m \cdot \tau_{\text{abs}}(\lambda)), \tag{6}$$

where $A(\lambda)$ is the surface albedo, $F_0(\lambda)$ the solar irradiance, $\tau_{\text{abs}}(\lambda)$ the absorption optical depth, and $m = \cos(\theta_0)^{-1} + \cos(\theta_v)^{-1}$ is the air mass factor, which scales the effective path length according to the solar zenith angle θ_0 and the viewing angle θ_v. To further simplify, we only account for molecular absorption, which makes the non-scattering model completely independent of aerosols. As before, non-scattering radiances $\tilde{I}(\lambda)$ are convolved with the ISRF to obtain simulated radiances $\tilde{I}_k$ for each instrument channel k. Due to the simplicity of this model, Jacobians can be computed analytically, requiring only a single forward simulation.

The non-scattering approximation captures the dominant structure of spectrometer measurements, including molecular absorption features, the spectral continuum shape, and instrument characteristics, while avoiding expensive computations associated with simulating scattering events. However, by neglecting the modification of the effective light path by aerosol scattering, it can introduce large systematic errors in retrieved gas concentrations, particularly under high aerosol loads (Butz et al., 2009).

2.5 Neural network-based emulators

To enable both fast and accurate retrievals, we replace the computationally expensive full-physics forward model and its Jacobian with neural network-based emulators. These emulators are trained to approximate the physical input–output mapping $\mathbf{I} = F(\mathbf{x}, \mathbf{b})$, and the corresponding Jacobians are obtained by differentiating the trained emulators. Neural networks are well suited for this task given their ability to model complex non-linear functions with high-dimensional outputs, and have been shown to outperform other machine learning approaches for spectrometer forward model emulation (Brence et al., 2023). In addition, they can exploit GPU acceleration for efficient training and fast predictions. We consider two types of emulators: a fully data-driven end-to-end emulator that directly approximates the full forward model, and two hybrid emulators that combine neural networks with non-scattering radiative transfer simulations as a physics-based backbone. Figure 1 illustrates the information flow and network architectures of these three emulator variants. All emulators use separate neural networks for each spectrometer band, allowing each network to focus on the spectral features relevant to its wavelength range.

2.5.1 Base emulator

The end-to-end emulator, referred to hereafter as the base emulator, approximates the full physical forward model and relies solely on the input-output examples in the training data. We implement this as a multi-layer perceptron (MLP), consisting of a sequence of fully-connected layers. Each layer applies an affine transformation with learnable weights and biases followed by a nonlinear activation function. In our setup, the MLP consists of three components: an encoder layer that maps forward model

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

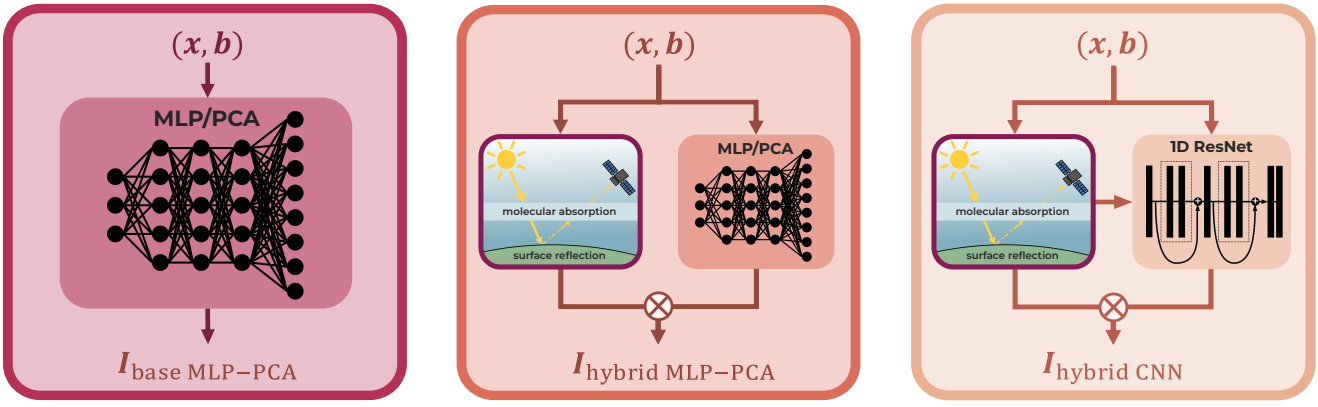

Figure 1. Overview of the three neural network-based forward models developed in this study. The base emulator (left) uses a combination of multi-layer perceptron (MLP) and principal component analysis (PCA) to predict radiance spectra directly from forward model inputs $(\mathbf{x}, \mathbf{b})$. The other two emulators use a hybrid approach combining non-scattering radiative transfer simulations with a scattering correction that is learned either by the same MLP-PCA network as the base emulator (center), or by a 1-dimensional convolutional ResNet architecture taking as input the non-scattering radiance spectrum (right).

inputs to a latent representation, a core consisting of several hidden layers of equal width that process this representation, and a decoder that outputs predictions $\mathbf{z}_{\text{base-NN}} \in \mathbb{R}^K$ across all K wavelength channels in the band.

Since spectrometer measurements typically span several orders of magnitude across wavelengths and atmospheric conditions, we apply a logarithmic transformation prior to neural network training. This emphasizes relative rather than absolute errors and improves numerical stability during training. At prediction time, the neural network outputs $\mathbf{z}_{\text{base-NN}}$ are mapped back to physical radiance space by inverting this transformation:

$$\mathbf{I}_{\text{base-NN}} = \exp(\mathbf{z}_{\text{base-NN}}), \tag{7}$$

which yields strictly non-negative final emulator predictions, as required by basic physical principles.

In addition to their large dynamic range, spectrometer measurements pose a challenge for data-driven emulators due to their high dimensionality. Typical spectra comprise hundreds or even thousands of spectral pixels, capturing fine-scale absorption features across large wavelength ranges. Learning such high-dimensional structures directly can be difficult, particularly when training data is limited. To simplify the learning problem and avoid overfitting, dimensionality reduction techniques, such as principal component analysis (PCA), are commonly applied to compress the measurements into a smaller set of components (Gómez-Dans et al., 2016; Verrelst et al., 2017; Brence et al., 2023; Lamminpää et al., 2025). The machine learning emulator can then be trained to predict these reduced components rather than the full measurement vectors.

Here, we adopt a similar approach, but integrate the PCA directly into the neural network architecture. Prior to training, a PCA with $P \ll K$ components is fitted to the training targets $\mathbf{z}$. The resulting PCA basis matrix $\mathbf{U} \in \mathbb{R}^{K \times P}$ spans a low-dimensional spectral subspace capturing the directions of largest variance. Instead of training directly in this PCA space, we

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

incorporate the inverse PCA transformation $\mathbf{z} = \mathbf{U}\mathbf{c}$ into the MLP decoder. That means, the decoder consists of a learnable layer that predicts PCA coefficients $\mathbf{c} \in \mathbb{R}^P$, followed by a linear layer with fixed weights that are initialized to $\mathbf{U}$.

This design constrains predictions to a structured low-dimensional subspace while retaining optimization in the full spectral domain. Although this slightly increases training cost, it reduces the risk that small approximation errors in PCA coefficient space translate into disproportionately large errors in spectral space, thereby improving robustness for downstream retrieval applications.

2.5.2 Hybrid emulators

Purely data-driven emulators can in principle achieve high accuracy given large enough training data and network capacity (Hornik et al., 1989). However, if either is insufficient, they may fail to resolve fine-scale absorption features or produce systematic prediction errors, both of which can degrade retrieval accuracy. Moreover, without any grounding in physical knowledge, they may extrapolate poorly to atmospheric conditions outside the training domain. The hybrid emulators developed here aim to mitigate these limitations by incorporating physical knowledge in the form of non-scattering radiances $\tilde{\mathbf{I}}$ (see Section 2.4).

To account for scattering-induced errors, we use neural networks to predict multiplicative correction factors, which are applied to the non-scattering spectra to recover full-physics radiances. Rather than learning all relevant physics from scratch, this allows the neural networks to focus specifically on learning scattering-induced features, while relying on the non-scattering approximation to provide information on the continuum and molecular absorption features. Similar to the base emulator, the correction factors are modeled in logarithmic space, which converts multiplicative errors into additive ones, improves numerical stability, and ensures that final emulator predictions remain strictly non-negative:

$$\mathbf{I}_{\text{hybrid-NN}} = \tilde{\mathbf{I}} \odot \exp\left(\mathbf{z}_{\text{hybrid-NN}}\right), \tag{8}$$

where $\odot$ denotes element-wise multiplication.

We consider two different network architectures to predict these log-transformed correction factors: an MLP and a convolutional neural network (CNN). The MLP variant closely follows the base emulator setup, with an encoder, a core of hidden layers, and a decoder that predicts the principal components of the hybrid training targets before reconstructing them for all spectral channels via a fixed inverse PCA layer.

In contrast, the CNN variant operates directly in spectral space, and thus respects the inherent physical structure of spectrometer measurements. To this end, the non-scattering approximation $\tilde{\mathbf{I}}$ is used as an additional, physically informed, and spectrally structured input on which the network operates. The non-scattering radiances are log-transformed, concatenated with the forward model inputs at each wavelength channel, and processed by a small MLP to extract relevant spectral features. To provide the network with information about absolute spectral positions, a sinusoidal positional encoding vector is added to the feature vector in each channel.

The resulting spectral representation $\mathbf{H} \in \mathbb{R}^{K \times M_{\text{in}}}$, consisting of M_{in} features for each channel k, is processed by a lightweight one-dimensional ResNet (He et al., 2016). The ResNet architecture consists of several blocks that combine convolutional layers with residual connections to facilitate easier optimization and deeper networks. Throughout the network, the number of

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

feature channels in the convolutional layers is kept constant at M_{in}, preserving the dimensionality of the spectral representation. Only in the final convolutional layer is the number of channels reduced to $M_{\text{out}} = M_{\text{in}}/2$, effectively compressing the spectral features and encouraging the network to retain only the most relevant information for the final prediction. To preserve the sequence length after convolution while avoiding boundary artifacts at the edges of the spectrum, replicate padding is applied before all convolutional layers. Importantly, the convolutional structure enables the network to exploit the sequential nature of spectrometer measurements, where nearby wavelength channels are strongly correlated due to spectral broadening and mixing introduced by the ISRF. Finally, a small MLP maps each spectral feature vector $f_{\text{1D-ResNet}}(\mathbf{H})_k \in \mathbb{R}^{M_{\text{out}}}$ to a single log-transformed correction factor.

2.5.3 Emulator Jacobians

Assuming that the emulators learn a locally accurate and smooth approximation of the full-physics forward model, the corresponding emulator Jacobians can be obtained by differentiating the trained neural network-based forward models with respect to each state vector parameter $\mathbf{x}_j$.

For the hybrid emulators, these derivatives decompose according to the product rule:

$$\frac{\partial \mathbf{I}_{\text{hybrid-NN}}}{\partial \mathbf{x}_j} = \frac{\partial \tilde{\mathbf{I}}}{\partial \mathbf{x}_j} \odot \exp(\mathbf{z}_{\text{hybrid-NN}}) + \tilde{\mathbf{I}} \odot \frac{\partial \exp(\mathbf{z}_{\text{hybrid-NN}})}{\partial \mathbf{x}_j}, \tag{9}$$

separating sensitivity to the non-scattering backbone from sensitivity to the learned scattering correction. Since albedo and gas parameters primarily affect the radiance through molecular absorption and surface reflection captured by the non-scattering approximation, their derivatives are dominated by the first term, whereas for aerosol parameters the scattering contribution and thus the second term dominates. In both cases, the accuracy of hybrid emulator derivatives can benefit from the non-scattering backbone —either directly via the physics-based sensitivity to albedo and gas parameters, or indirectly through the reduced complexity of the learning problem that helps the neural networks learn a smoother and more accurate local approximation, which in turn results in more reliable gradients.

2.6 Data

2.6.1 Synthetic training and test datasets

To train and evaluate the neural network emulators, we generate random ensembles of synthetic scenes for which S5/UVNS radiance spectra are simulated with the full-physics forward model defined in Section 2.3 as well as the non-scattering forward model defined in Eq. (6). For each scene, satellite geometry, surface albedo, aerosol and gas parameters are randomly sampled from uniform distributions with predefined ranges. The reference profiles of O_2, CO, CO_2, CH_4, and H_2O are defined based on the AFGL US standard atmosphere (Anderson et al., 1986). For CO_2 and CH_4, the AFGL profiles are adjusted to match current background concentrations, of approximately 430 ppm and 1900 ppb, respectively. For CO_2, CH_4, and H_2O, the total column concentration is then varied by sampling different profile scaling factors, while O_2 and CO are kept fixed. Surface elevation,

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

air pressure and temperature profiles are sampled jointly from global land-only ERA5 reanalysis data (Hersbach et al., 2020) covering 4 representative days (10 Feburary, 10 May, 10 August and 10 November) of the years 2019-2022.

Two separate ensembles are generated with different parameter ranges (see Table 2). The first ensemble, consisting of 10^6 scenes, is used for training and initial evaluation of radiance predictions and is designed to cover broad parameter ranges in order to develop robust and versatile emulators. 85% of this dataset is used for training, 5% for validation (i.e. early stopping and hyperparameter tuning) and 10% for testing (i.e. final radiance evaluation). The second ensemble, consisting of 10^3 scenes, is reserved exclusively for the final evaluation of emulator-based retrievals and focuses on more realistic parameter ranges representative of actual satellite observations. For all scenes in this retrieval test dataset, we add instrument noise according to Eq. (1) to obtain realistic measurements.

Additionally, we create several variants of the retrieval test dataset to assess the ability of the different emulators to generalize to high-emission scenarios where the CH_4 scaling factor exceeds the range covered by the training ensemble. For each variant, the CH_4 scaling factor is fixed to a prescribed value, while all other parameters remain identical to those in the original retrieval test dataset to ensure comparability. We include two variants falling within the training range, two at the upper and lower boundary of the training range, and three variants exceeding the training range with column-averaged XCH_4 reaching up to 2650 ppb.

2.6.2 Neural network inputs

Table 3 lists the variables provided as input to the neural networks. These variables are derived from the forward model inputs $(\mathbf{x}, \mathbf{b})$. Constants such as aerosol refractive indices, gas profile shapes, solar irradiance, and instrument characteristics, which do not vary in the training data, are not provided to the neural networks but are implicitly encoded by the trained network weights.

Temperature profiles are reduced from 20 pressure layers to two latent variables using a PCA, which is fitted to the training data prior to neural network training. This captures the dominant modes of variability in atmospheric temperature profiles while limiting the dimensionality of the neural network inputs, thus reducing the risk of overfitting.

All input variables are standardized using the mean and standard deviation computed from the training data before being provided to the neural networks. This standardization is applied after any logarithmic or cosine transformations of the inputs. Temperature profiles are standardized per vertical layer such that each vertical layer has zero mean and unit variance before fitting and applying the PCA. Log-transformed non-scattering radiances provided to the hybrid CNN emulator are standardized per spectral band to preserve correlations between neighboring channels.

2.7 Neural network training

2.7.1 Training procedure

The training dataset described in Section 2.6 consists of representative input-output examples $\{((\mathbf{x}_n, \mathbf{b}_n), \mathbf{I}_n, \tilde{\mathbf{I}}_n)\}_{n=1}^{N}$. Prior to training, full-physics radiances $\mathbf{I}_n$ are transformed into training targets $\mathbf{z}_n$ based on Eq. (7) and (8), respectively. For numer-

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

Table 2. Parameter ranges for the two random ensembles used for training, validation and initial radiance evaluation (train/val/test), and for final retrieval-based evaluation respectively. The vertical profiles of trace gases are assumed to follow the profiles of the AFGL US standard atmosphere, only adjusted by a scaling factor.

category	**parameters**	**sampling range**	
		train/val/test	**final evaluation**
geometry	solar zenith angle	$[0°, 75°]$	$[0°, 70°]$
	viewing zenith angle	$[0°, 35°]$	$[0°, 30°]$
	solar azimuth angle	$[-180°, 180°]$	$[-180°, 180°]$
	viewing azimuth angle	$[-180°, 180°]$	$[-180°, 180°]$
gases	H_2O scaling factor	$[0.85, 1.15]$	$[0.9, 1.1]$
	CH_4 scaling factor	$[0.85, 1.15]$	$[0.9, 1.1]$
	CO_2 scaling factor	$[0.85, 1.15]$	$[0.9, 1.1]$
aerosol	AOD at 550nm	$[0.0001, 1.0]$	$[0.001, 0.3]$
	central height z_{aer}	$[0\text{ km}, 10\text{ km}]$	$[0\text{ km}, 8\text{ km}]$
	size parameter p	$[0.1, 6.0]$	$[3.0, 5.0]$
surface	albedo (NIR)	$[0.001, 1.0]$	$[0.02, 0.8]$
	albedo (SWIR1)	$[0.001, 1.0]$	$[0.02, 0.8]$
	albedo (SWIR3)	$[0.001, 1.0]$	$[0.02, 0.8]$
	elevation z_{surf}	ERA5 (2019-2021)	ERA5 (2022)
temperature	vertical profile	ERA5 (2019-2021)	ERA5 (2022)
air pressure	vertical profile	ERA5 (2019-2021)	ERA5 (2022)

ical stability, these training targets are standardized to $\hat{\mathbf{z}}_n = (\mathbf{z}_n - \boldsymbol{\mu})/\boldsymbol{\sigma}$, where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are vectors of means and standard deviations computed from the training dataset. For the MLP-based networks, elements μ_k and σ_k are computed separately for each channel k, while for CNN-based networks a single mean and standard deviation is computed per spectral band to preserve correlations between neighboring channels. This standardization is inverted before transforming training targets into final radiances at prediction time.

During training, the learnable parameters $\boldsymbol{\eta}$ of the neural network function f_{NN} are adjusted iteratively to minimize the mean-squared error loss function

$$\mathcal{L}_{MSE} = \frac{1}{N \cdot K} \sum_{n=1}^{N} \sum_{k=1}^{K} \left(f_{\text{NN}}(\mathbf{x}_n, \mathbf{b}_n; \boldsymbol{\eta})_k - \hat{z}_{n,k} \right)^2 . \tag{10}$$

The optimization is performed using the Adam optimizer (Kingma and Ba, 2015), with a dynamic learning rate that first increases linearly from 10^{-8} to 10^{-3} over the first 7500 iterations (warm-up phase, scaled by the fraction of training data used)

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

Table 3. Forward model parameters used as emulator inputs per spectral band, together with their respective units and transformations applied before feeding them to the neural networks. For temperature profiles, only the first two principal components are used.

variable	unit	transformation	band
solar zenith angle	radians	$\cos(\cdot)$	all
viewing zenith angle	radians	$\cos(\cdot)$	all
relative azimuth angle	radians	—	all
scattering angle	radians	—	all
surface pressure	hPa	—	all
average albedo across band	—	$\log(\cdot)$	all
aerosol size parameter	—	—	all
aerosol optical depth at 550nm	—	—	all
aerosol central height	m	—	all
total column density of H_2O	molecules/m^2	—	SWIR1, SWIR3
total column density of CO_2	molecules/m^2	—	SWIR1
total column density of CH_4	molecules/m^2	—	SWIR1, SWIR3
temperature profile	K	PCA	all

and then decays to 10^{-5} using cosine annealing. Training is carried out with a batch size of 256 for a maximum of 75 epochs, with early stopping applied if the validation loss does not improve for 20 consecutive epochs. To ensure stable optimization, gradient values are clipped to a maximum absolute magnitude of 1.0 during backpropagation. To reduce overfitting, we apply a small dropout rate of 0.001.

All neural networks are implemented using PyTorch (Paszke et al., 2019), and take 1-2 hours to train on two NVIDIA A100 GPUs.

2.7.2 Hyperparameter selection

The number of principal components P used by the MLP-based emulators is determined separately for each band by repeatedly fitting the PCA with increasing P to the standardized training targets. For each P, the corresponding reconstruction error is evaluated using the validation set. The reconstruction error is quantified as the root-mean-square error (RMSE) between original radiances $\mathbf{I}$ and reconstructed radiances $\mathbf{I}_{\mathrm{PCA}} = \exp(\mathbf{U}\mathbf{c} \odot \boldsymbol{\sigma} + \boldsymbol{\mu})$, relative to the continuum (approximated by $\max(\mathbf{I})$). The final dimensionality P is chosen as the smallest number of components that satisfies two criteria: (i) the relative reconstruction error falls below 0.001% and (ii) further increasing P reduces the reconstruction error by less than 15%. This ensures that the retained components provide near-lossless spectral reconstruction while avoiding unnecessarily large latent dimensionality. Following this procedure, we use 100 components in the NIR band, 160 in SWIR1 and 140 in SWIR3. For all three bands, the selected components explain more than 99.99999% of the variance in the training data.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

To determine the optimal architecture for the base emulator, we performed a grid search over the width and number of hidden layers of the MLP core. For each combination, separate emulators were trained and their predictive performance was evaluated on the validation set. We generally observe that the relative RMSE decreases with the layer width, but at the expense of an increase in the relative bias. Using 3 hidden layers of width 128 seems to offer a good trade-off across spectral bands. To ensure a fair and consistent comparison between emulator variants, the same architecture is used for the MLP-based hybrid emulator.

For the CNN-based hybrid emulator, we performed a separate grid search over the number of ResNet blocks, the convolutional kernel size, the spectral feature dimension, and the number of layers in the MLPs used for input processing and final prediction. As before, the architecture yielding the best trade-off between relative RMSE and bias on the validation set is selected. The final architecture consists of two ResNet blocks with a convolutional kernel size of 5, feature dimension $M_{\text{in}} = 128$, and input/output MLPs with one hidden layer each.

Based on additional evaluations on the validation set, we choose to use SiLU activation functions (Elfwing et al., 2018) for all neural network emulators.

2.8 Retrieval setup

To assess how well the emulators perform when embedded within the retrieval algorithm, we run retrievals on the simulated Sentinel-5/UVNS measurements in the retrieval test set, which are consistent with the forward model used to generate the training data, up to instrument noise. This setup allows us to directly compare retrieved parameters with their true values and to systematically evaluate how emulator-induced forward model errors propagate to retrieval errors.

We perform multi-band retrievals, in which radiance measurements from all spectrometer bands are fitted simultaneously. The retrieved state vector $\mathbf{x}$ consists of the albedo in each spectral band, scaling factors for H_2O, CO_2, and CH_4, and the aerosol parameters p, N_{aer}, and z_{aer}. The retrieval is initialized based on the a priori state vector $\mathbf{x}_a$, using fixed first guesses for aerosol parameters ($p = 4.0, \text{AOD}_{550} = 0.1, z_{aer} = 5000$m) and surface albedo (0.1 in each spectral band). The a priori vertical profiles of trace gases are obtained by randomly scaling the true profile by up to $\pm 10\%$. Following the operational setup for TROPOMI CH_4 retrievals (Hu et al., 2016), we define the diagonal weighting matrix (see Section 2.2) as $W_{ii} = 1/x_{a,i}$, which ensures that all parameters are of the same order of magnitude and contribute equally to the side constraint. A maximum of 10 iterations is used, and the retrieval is considered converged when the relative improvement in the goodness of fit (reduced χ^2) falls below 0.01, or the absolute improvement falls below 0.05.

2.8.1 Emulator-retrieval integration

While the retrieval algorithm is implemented in Fortran, the neural network emulators are trained in Python using PyTorch. To make the two compatible, we export the trained networks as standalone executables via TorchScript, which can subsequently be loaded and called within Fortran-based retrieval codes via the FTorch interface (Atkinson et al., 2025).

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

2.8.2 Emulator Jacobian computation

As discussed in Section 2.5.3, we obtain emulator Jacobians via differentiation of the trained neural network forward models. While this can in principle be done analytically using PyTorch's automatic differentiation (Paszke et al., 2017), this functionality is currently not available when the networks are executed via FTorch from Fortran. Therefore, we approximate the required derivatives using finite differences. The step sizes are determined by comparing against analytical derivatives computed in PyTorch, ensuring that the numerical approximation errors are negligible. In terms of accuracy, this implementation is thus representative of automatic differentiation, while introducing some computational overhead due to repeated forward passes through the emulator.

3 Results

We evaluate our neural network emulators in terms of three aspects: spectral accuracy of the predicted radiances, consistency of the associated Jacobians with those of the full-physics forward model, and retrieval performance compared to physics-based retrievals. To assess the value of incorporating physical knowledge into the emulator design, we place a specific focus on the sensitivity to training dataset size and the ability to generalize to high-emission scenarios beyond the training distribution.

3.1 Spectral evaluation

3.1.1 Radiance residuals

To determine the spectral accuracy of the trained neural network emulators, the predicted radiances are compared to the corresponding target radiances in the test set withheld from training (see Section 2.6). Figure 2 shows the distribution of spectral residuals, normalized by the continuum (approximated by the maximum radiance in each band) for each emulator. Overall, all emulators achieve low spectral biases (≤ 0.11 % compared to $\gg 2$ % for non-scattering approximations) and reasonable continuum-normalized RMSE (≤ 2.03 % compared to $\gg 13$ % for non-scattering approximations) across all spectral bands, indicating that both end-to-end and hybrid approaches can reproduce the overall shape and structure of the simulated radiance spectra. However, the hybrid emulators achieve substantially lower errors, with RMSE ranging from 0.14 % to 0.17 % across spectral bands compared to 0.62-2.03% for the base emulator, and result in more stable residuals across instrument channels. Spectral biases are also smaller than for the base emulator. The two hybrid emulators show very similar spectral accuracy, with the hybrid CNN variant achieving slightly lower RMSE in the SWIR1 and SWIR3 bands.

3.1.2 Impact of training data size

In addition, we investigate the sensitivity of each emulator type to the number of input-output examples in the training dataset. Figure 3 shows the RMSE and absolute bias of continuum-normalized radiance residuals as a function of the number of training examples, which are randomly sampled from the entire training dataset. Overall, both hybrid emulator types consistently require less training data to achieve the same spectral accuracy as the base emulator, confirming our hypothesis that incorporating

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

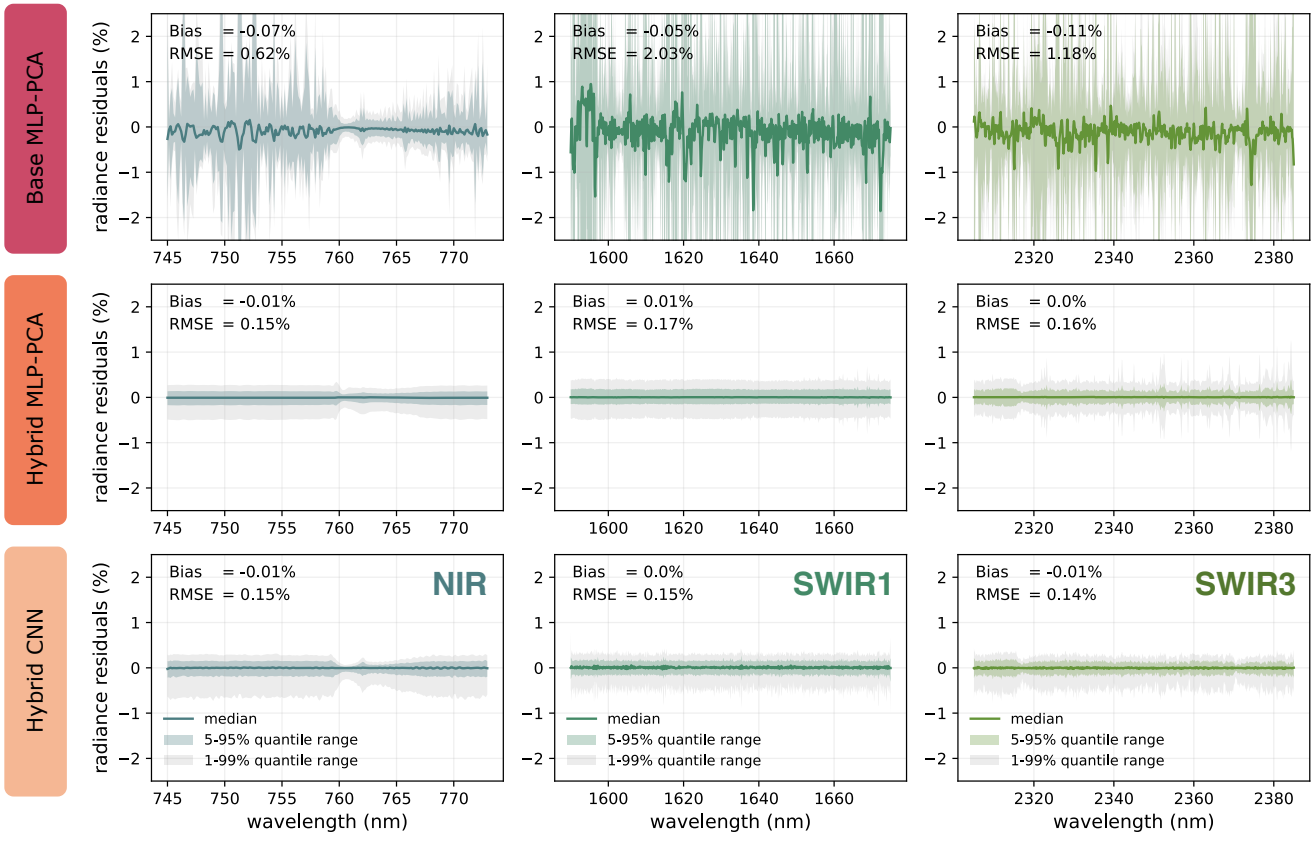

Figure 2. Spectral accuracy of the trained neural network emulators evaluated based on 10^5 synthetic measurements of the Sentinel-5/UVNS spectrometer. All radiance errors are calculated relative to the continuum of the corresponding physics-based reference spectrum.

physical knowledge improves data efficiency. However, the two metrics show distinct patterns in how the three emulator variants respond to data limitations.

In terms of RMSE, three distinct regimes can be identified across the training dataset sizes considered. In the intermediate data regime (4×10^3 to 10^5 scenes), all emulators exhibit a similar dependence of RMSE on training data size, with the base emulator requiring approximately 2-3 times more training data than the hybrid variants to achieve the same accuracy. In the low-data regime ($< 3 \times 10^3$), the base emulator's RMSE increases much more steeply with decreasing data, while the hybrid MLP-PCA maintains a similar rate as in the intermediate regime. The hybrid CNN is the most data-efficient of the three emulator variants, with its RMSE remaining below 10 % for all considered dataset sizes. In the high-data regime ($> 10^5$), the two hybrid emulators continue to improve steadily as more training examples are added, suggesting that their spectral accuracy could improve even further with larger training dataset size. In contrast, the base emulator accuracy plateaus, indicating that its network capacity has been saturated and further improvements would require larger network architectures.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

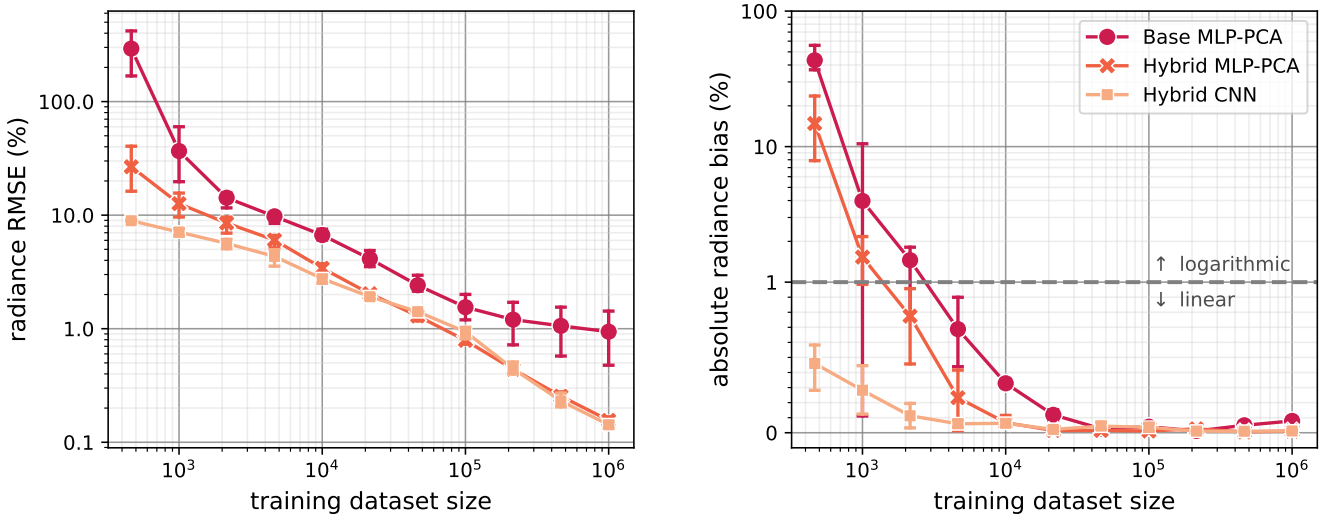

Figure 3. RMSE and bias of continuum-normalized radiance residuals as a function of the number of training samples. Markers and error bars show the mean and standard deviation across spectral bands. Note that the radiance bias is plotted on a symmetric log-scale with values between -0.2% and 0.2% spaced linearly, while larger magnitudes are spaced logarithmically.

Spectral biases, on the other hand, show a clear threshold effect. For large enough datasets, all emulators achieve small absolute biases ≤ 0.1 %. But as the number of training samples decreases, a critical point is reached at which the bias magnitude starts to increase sharply. Importantly, this critical point depends on the emulator design: it occurs earliest for the base emulators and latest for the hybrid CNN emulator, which maintains low biases ≤ 0.1 % with as little as 4×10^3 samples and remains below 1 % for all considered dataset sizes.

The data efficiency of the hybrid CNN emulator can be attributed to the inductive bias introduced by the convolutional architecture, which naturally captures the sequential structure and local spectral correlations in spectrometer measurements. By encoding this structure directly into the network architecture, the hybrid CNN can learn a more compact and generalizable representation of the scattering correction, reducing its dependence on large amounts of training data.

3.1.3 Jacobian evaluation

Beyond spectral precision, accurate Jacobians are essential for gradient-based retrieval algorithms (see Section 2.2). We therefore evaluate how well the emulator-based Jacobians match those of the full-physics forward model. Importantly, the neural networks are not explicitly trained to reproduce physical gradients, which means that Jacobian accuracy is an emergent property of the learned input-output mapping. To evaluate this in a realistic retrieval setting, we run full-physics retrievals while comparing emulator-based and full-physics Jacobians at each retrieval iteration.

Fig. A1-A3 in Appendix A show that all neural network-based emulators are able to capture the direction and overall magnitude of the true physical gradients. Generally, the accuracy of emulator derivatives is related to the true instrument sensitivity:

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

aerosol central height exhibits both the lowest sensitivity and the lowest derivative alignment, while albedo parameters exhibit the highest sensitivity and best derivative alignment, across all emulator types. Similarly to the radiance accuracies, the two hybrid emulators provide more accurate Jacobians than the base emulator. As expected from the decomposition discussed in Section 2.5.3, albedo and gas derivatives show near-perfect correlation with the full-physics reference ($R > 0.99$). Notably, the aerosol-related derivatives, which are dominated by the learned neural network correction, are also improved compared to the base emulator, consistent with the hypothesis that the reduced complexity of the learning problem leads to smoother and more accurate local approximation.

3.2 Retrieval evaluation

Finally, we assess the effects of radiance and Jacobian errors on the final retrieval performance. To this end, we run retrievals using the neural network-based emulators as forward models and compare the retrieved XCH_4 and XCO_2 to both full-physics and non-scattering retrievals. The full-physics retrievals are based on the same forward model used to generate the training and test data. This provides a reference for the best achievable retrieval performance and allows us to quantify the impact of replacing the forward model with neural network-based emulators. Non-scattering retrievals, on the other hand, represent a lower-bound or worst-case scenario baseline that the hybrid emulators should outperform to be useful in practice.

3.2.1 Overall performance

Figure 4a shows retrieved vs. simulated XCH_4 and XCO_2, for physics-based and emulator-based retrievals respectively. We find that all emulator types consistently improve upon non-scattering retrievals, with the hybrid emulators outperforming the base emulator in terms of both retrieval precision and accuracy. Relative to the full-physics reference, the base emulator introduces additional retrieval errors of 7.22 ppb for XCH_4 and 4.25 ppm for XCO_2, and increases the bias in XCO_2 by 1.21 ppm. In contrast, both hybrid emulators reduce emulator-induced retrieval errors to less than 1.2 ppb for XCH_4 and less than 0.5 ppm for XCO_2, without increasing retrieval biases. In line with our spectral analysis, the hybrid CNN variant achieves slightly lower errors and biases for both XCH_4 and XCO_2. However, the impact of neural network architecture design is small compared to the substantial improvements of the hybrid approach over the base emulator more generally.

Similar performance patterns are observed for the retrieved aerosol parameters, with the hybrid emulators providing more accurate estimates than the base emulator across all aerosol properties (see Appendix B).

The small retrieval errors of the hybrid emulators are also reflected in the reduced χ^2 statistic (see Figure 4b), with values close to unity confirming that their residuals are well-behaved across all spectral channels, including deep absorption features with low measurement noise. The base emulator, on the other hand, exhibits large $\chi^2 \gg 10$, even exceeding that of the non-scattering model despite achieving better retrieval accuracy. This illustrates that an elevated χ^2 does not necessarily indicate poor retrieval quality. Systematic spectral errors can be strongly amplified by the noise weighting in the χ^2 statistic, even when the emulator captures sufficient information on gas absorption and aerosol scattering to produce meaningful retrievals. This should be taken into account when applying standard χ^2-based quality filters to emulator-based retrievals, as overly strict thresholds may inadvertently reject retrievals of acceptable quality.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

3.2.2 Computational efficiency

Zooming in on computational efficiency, we estimate speed-ups relative to the full-physics forward model based on total CPU runtimes for all three spectral bands. We distinguish between (i) forward model evaluations alone, (ii) the computational cost of a single retrieval iteration (including both forward model and Jacobian computations), and (iii) the expected computational cost of a complete retrieval, which additionally accounts for differences in the average number of iterations until convergence. All runtimes are measured on a single core of an AMD EPYC 7543 processor (32 cores, 3.7 GHz) and averaged over multiple simulations.

We find that the base emulator is by far the fastest model, achieving a speed-up of factor 678.3 for forward model evaluations. The non-scattering forward model is 18 times faster than the full-physics model, which defines the maximum speed-up attainable by the hybrid emulators. The hybrid MLP-PCA variant comes close to this limit, with the neural network correction adding only negligible computational overhead, resulting in an overall speed-up of a factor of 17.1. The hybrid CNN variant achieves a lower speed-up of factor 9.3 due to the larger size and complexity of the convolutional network architecture.

Including Jacobian computations affects the speed-ups differently across emulator types. Since the non-scattering model benefits from analytical Jacobian computations, it scales more favorably than the full-physics forward-adjoint approach, resulting in an increased speed-up of 32.2. This also translates to the hybrid MLP-PCA emulator, whose runtime is dominated by the non-scattering backbone such that the additional overhead of Jacobian calculations for the neural network component is negligible, resulting in an increased speed-up of 29.8. For the base emulator and hybrid CNN, however, the neural network components contribute substantially to the overall runtime. And as the finite difference Jacobian computation requires one additional neural network evaluation per state vector parameter, this introduces significant overhead compared to the full-physics forward-adjoint approach, resulting in decreased speed-ups of 340.0 and 6.4 respectively.

Interestingly, all emulator types require fewer iterations to converge than full-physics retrievals, likely because retrieval residuals reach the spectral accuracy limit of the emulator, beyond which χ^2 improvements are minimal. Combining the per-iteration runtimes with this reduction in the average number of iterations yields expected speed-ups of the complete retrieval of 482.3 for the base emulator, 58.8 for the non-scattering model, 36.2 for the hybrid MLP-PCA, and 8.5 for the hybrid CNN.

Note that further acceleration of the neural network components could be achieved using GPU computation, and the overhead of Jacobian calculations could be substantially reduced by enabling automatic differentiation in FTorch, which would avoid the need for repeated forward passes when computing finite differences. In addition, we see potential in further improving the computational efficiency of the non-scattering model implementation, and thus the hybrid emulators.

3.2.3 Effects of aerosol optical depth

To further characterize retrieval performance, we investigate how XCH_4 and XCO_2 retrieval errors depend on aerosol optical depth (AOD) by grouping scenes in the retrieval test set into three different AOD ranges. As expected, non-scattering retrievals show quickly increasing errors and substantial biases with higher AOD due to neglected scattering effects (see Figure 5). In contrast, emulator-based retrievals are much less sensitive to AOD and both hybrid emulators consistently achieve accuracies

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

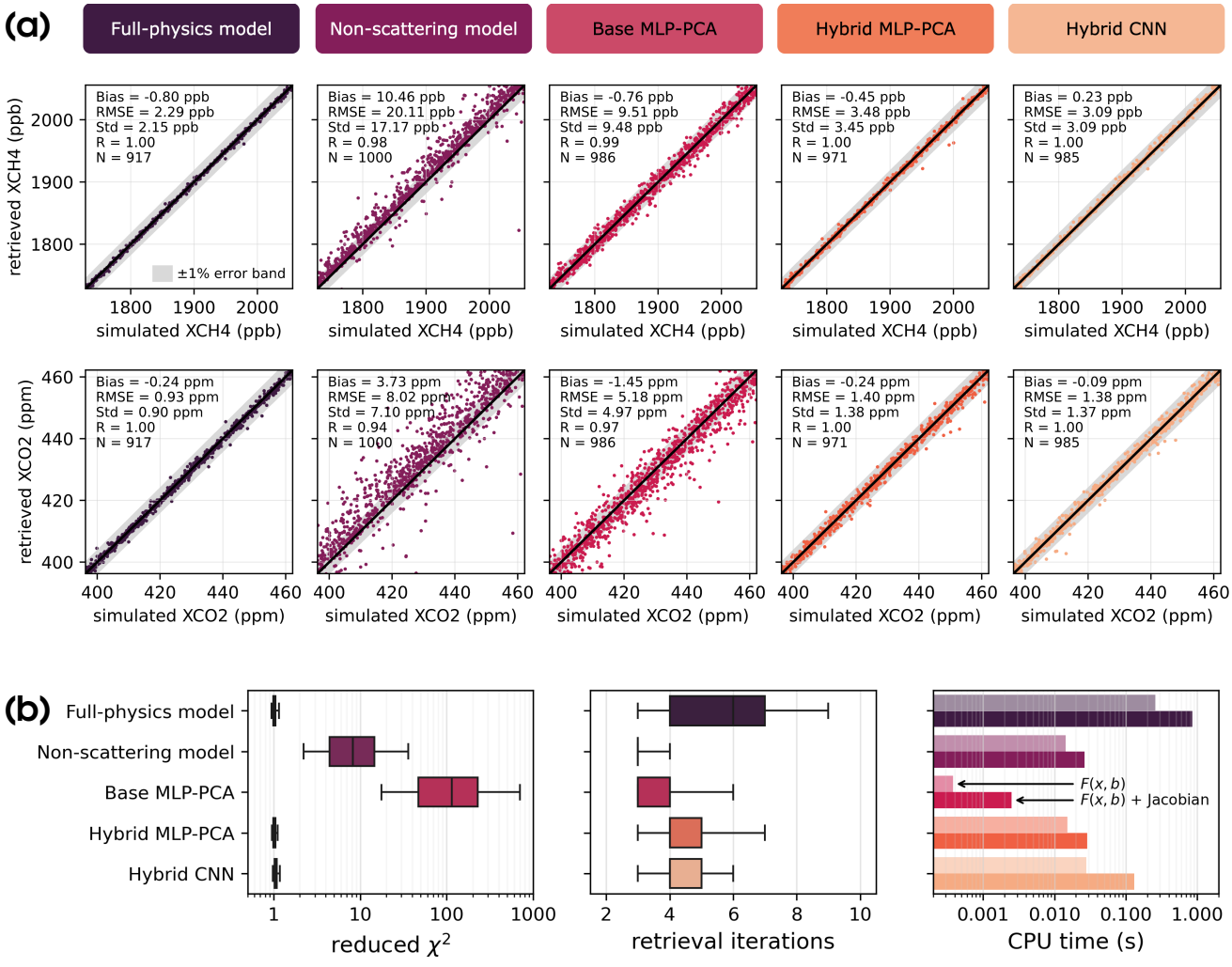

Figure 4. Retrieval evaluation for physics-based and neural network-based forward models. (a) Retrieved vs. simulated XCH_4 and XCO_2. (b) Additional retrieval diagnostics. From left to right: reduced χ^2 values measuring the goodness of fit of each retrieval, number of retrieval iterations until convergence, and CPU runtimes of the different forward models.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

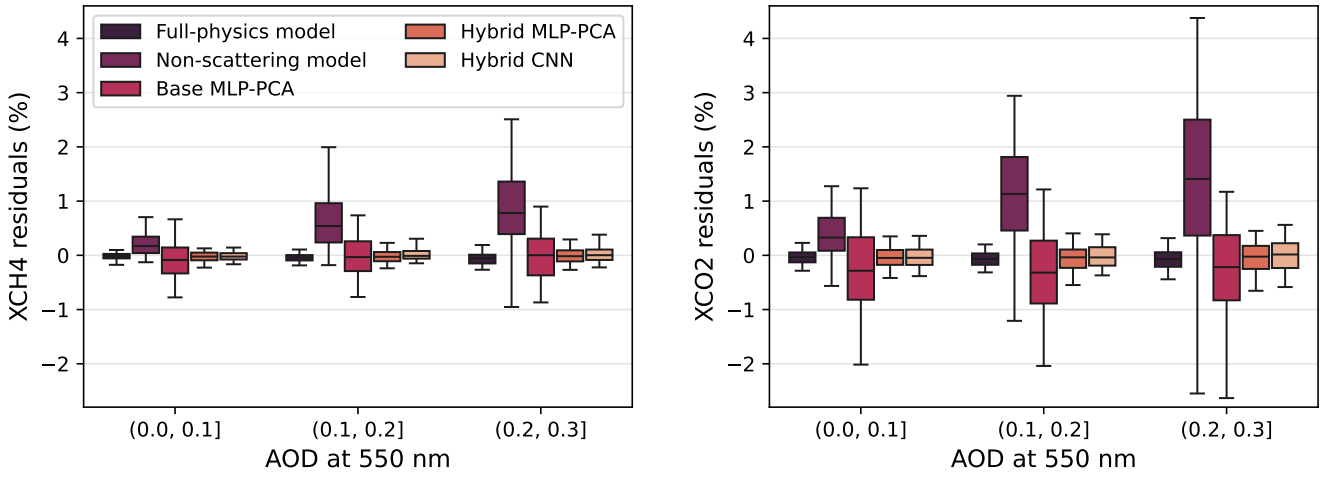

Figure 5. Distributions of XCH_4 (left) and XCO_2 (right) retrieval errors grouped by aerosol optical depth.

close to the full-physics reference, demonstrating that they accurately capture aerosol effects not accounted for by the non-scattering approximation.

3.2.4 Generalization to high CH_4 emission scenarios

Lastly, we evaluate retrievals for high-emission scenarios with gas concentrations beyond the training range to assess whether incorporating physical knowledge improves generalization to out-of-distribution conditions. Figure 6 shows how the precision and bias are affected by increasing XCH_4 concentrations. As expected, full-physics and non-scattering forward models, which build on well-established mechanistic understanding of radiative transfer and instrument physics, generalize to arbitrary gas concentrations and thus yield retrievals with consistent errors and biases. In contrast, the base emulator clearly suffers from extrapolation issues. As XCH_4 levels increase beyond the training range, the retrieval quality deteriorates quickly and eventually becomes worse than simple non-scattering retrievals. The hybrid emulators, on the other hand, can effectively leverage the underlying non-scattering approximation to generalize beyond the training range, resulting in only minor increases in retrieval errors and biases. As before, the hybrid CNN variant yields slightly more accurate and stable retrievals than the hybrid MLP-PCA emulator.

4 Discussion and conclusions

Accurate forward models are essential to obtain high-quality greenhouse gas estimates from spectrometer measurements, but present a major computational bottleneck in iterative retrieval algorithms. To alleviate this bottleneck, we have developed three different neural network-based forward model emulators and have explored their potential and limitations in the context of XCH_4 and XCO_2 retrievals from synthetic Sentinel-5 measurements. In the following, we discuss the implications of our

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

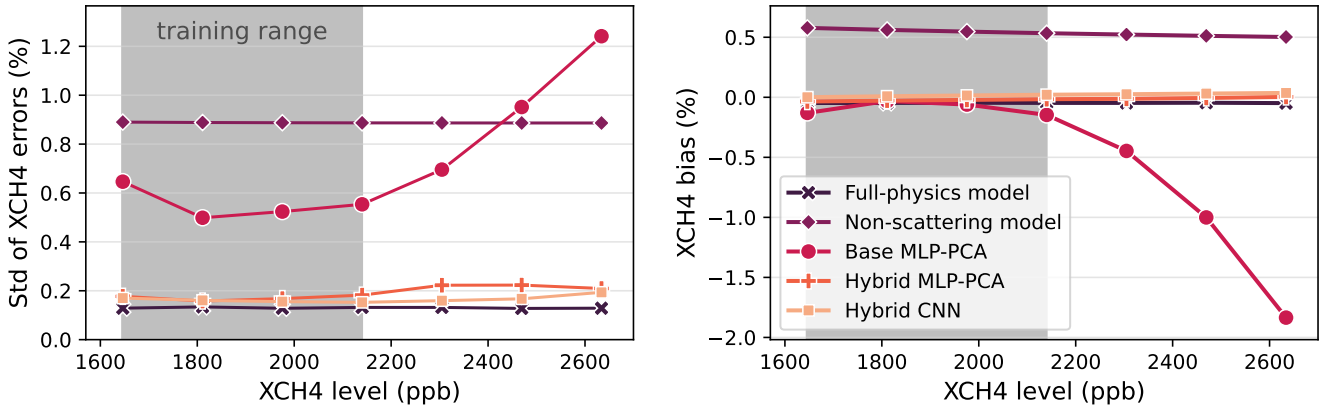

Figure 6. Robustness of XCH_4 retrievals to scenes with XCH_4 concentrations outside the training range, in terms of the standard deviation (left) and mean (right) of XCH_4 residuals.

results for emulator design choices (Section 4.1) and the application to operational Sentinel-5 retrievals (Section 4.2), and summarize our key findings and conclusions (Section 4.3).

4.1 Implications for emulator design

Our results show that, unlike previous studies on aerosol and surface retrievals (Bue et al., 2019; Gao et al., 2021; Nanda et al., 2019), purely data-driven end-to-end emulators struggle to meet the stringent requirements on XCH_4 and XCO_2 data products. We attribute this to the greater complexity of the greenhouse gas forward model mapping and the tighter accuracy requirements, which make it difficult for purely data-driven emulators to learn the relevant spectral features and effects in sufficient detail from training examples alone. In addition, the end-to-end approach fails to extrapolate reliably to atmospheric conditions outside the training distribution This is a practical concern for real-world applications, where the true atmospheric state is unknown and may deviate substantially from prior estimates, for example in the presence of strong anthropogenic emissions. Increasing training data coverage to mitigate this would quickly become infeasible given the computational cost of full-physics simulations. This highlights the need for more advanced hybrid approaches that incorporate prior knowledge on atmospheric physics to constrain the learning problem.

The hybrid approach proposed in this study achieves this by combining non-scattering radiative transfer simulations with a learned scattering correction, resulting in substantially improved spectral fidelity and retrieval performances much closer to the full-physics reference. Furthermore, being grounded in well-established radiative transfer physics, the non-scattering backbone remains informative across a wide range of atmospheric conditions and therefore helps the hybrid emulators to reliably extrapolate to gas concentrations far outside the training distribution. The reduced complexity of the learning problem also translates into substantially lower training data requirements, thus reducing the cost of emulator development. The saved computational budget could in turn be invested to increase the complexity and accuracy of the full-physics simulations used

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

for training data generation —for example by replacing the linear-k approximation with full line-by-line multiple scattering calculations, or adopting a vector rather than scalar radiative transfer model to account for polarization effects. This opens up opportunities for emulator-based retrievals that ultimately exceed the accuracy of operational full-physics retrievals that rely on several simplifications for computational tractability.

Nonetheless, these gains in accuracy, extrapolation capabilities, and data efficiency come at the cost of reduced speed-ups compared to the base emulator approach, as the overall runtime is limited by the physics-based non-scattering component. A similar trade-off exists between the two hybrid variants: the hybrid CNN achieves consistently better retrieval quality and data efficiency than the hybrid MLP-PCA, owing to its more expressive architecture and the inductive bias introduced by its convolutional layers. However, the increased complexity also results in substantially slower neural network evaluations, and thus roughly four times slower retrievals. The optimal emulator choice thus ultimately depends on the specific application, the associated accuracy requirements, and the computational resources available.

Furthermore, our results suggest that it is not necessary to train separate networks to replace the full-physics Jacobian, as proposed by Nanda et al. (2019) for example. Instead, we successfully obtain emulator Jacobians directly from the learned forward model mapping via differentiation, implying that our emulators learn a sufficiently accurate and smooth approximation of the full-physics forward model. This greatly simplifies emulator design as it avoids the development and maintenance of dedicated Jacobian networks. For more complex retrieval problems with additional aerosol parameters, a fine-tuning step in which the trained networks are further optimized to reduce errors in both radiances and their associated derivatives could further improve Jacobian accuracy as well as spectral fidelity (Luo et al., 2025).

4.2 Application to operational Sentinel-5 retrievals

The hybrid emulators developed in this study are promising candidates for implementation in operational Sentinel-5 data processing pipelines, with the hybrid MLP-PCA variant providing a particularly compelling balance between retrieval quality and speed. With emulator-induced retrieval errors as low as 0.06 % for XCH_4 and 0.11 % for XCO_2, this emulators can provide close-to full-physics retrieval quality remaining well within mission requirements, while being more than 30 times faster than retrievals using the (already optimized) full-physics forward model. Such speed-ups could substantially reduce the computational resources required for near-real-time processing of Sentinel-5 observations at the global scale, which has important implications for e.g. the Integrated Forecasting System (IFS) model used by the Copernicus Atmosphere Monitoring Service (CAMS).

Although the base emulator introduces considerably larger errors on XCH_4 (0.38 %) and XCO_2 (0.99 %), it still outperforms non-scattering retrievals and offers interesting opportunities in the context of Sentinel-5. With speed-ups of more than two orders of magnitude, it could for example be used to generate provisional or quick-look data products allowing for rapid hotspot detection. Alternatively, it could provide informative first guesses for full-physics retrievals, reducing the overall runtime by limiting the number of full-physics simulations required to reach convergence.

An important next step before implementing any of these operationally is to extend our evaluation to real satellite measurements. While the synthetic data setup adopted in this study enables a controlled analysis of emulator-induced errors in

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

isolation from other sources of uncertainty, applying emulator-based retrievals to real data introduces additional challenges, including instrument calibration errors and atmospheric processes not represented in the training data. Ensuring the robustness of emulator-based retrievals under these realistic conditions will be essential to pave the way for deployment in operational retrieval pipelines of Sentinel-5 and future missions such as CO2M. This will be addressed in a follow-up study.

4.3 Final conclusions

In summary, this study demonstrates that the development of accurate and fast neural network-based forward model emulators for greenhouse gas retrievals is challenging but becomes feasible when physical knowledge is incorporated into the emulator design through a hybrid approach. By combining fast non-scattering radiative transfer simulations with a learned scattering correction, we achieve very small emulator-induced retrieval errors that contribute only marginally to the overall error budget, while reducing the overall runtime by an order of magnitude. Compared to a purely data-driven end-to-end emulator, this hybrid approach is not only more accurate, but also requires less training data and generalizes more robustly beyond the training data.

More broadly, the emulator designs proposed here are readily transferable to retrievals of various other trace gases, as well as to other current and upcoming satellite missions such as GOSAT-GW and CO2M, whose increasing spatial resolutions lead to rapidly growing data volumes and hence greater computational demands. By shifting computational effort from operational retrievals to an offline training stage, such emulators can substantially reduce the resources required to process these large data volumes and thereby ensure that near-real-time global monitoring of the Earth's atmosphere and rapid detection of emission hotspots remain feasible as satellite missions continue to evolve.

Code and data availability. The Python code used to implement, train and evaluate our neural network models is available on Github at https://github.com/FionaLippert/Surrogate_RTM and an archived version is available at https://doi.org/10.5281/zenodo.21772755 (Lippert, 2026). The trained models, output data, and analysis code are available at https://doi.org/10.5281/zenodo.21792065 (Lippert et al., 2026a). The synthetic S5/UVNS measurements used for training and testing are available at https://doi.org/10.5281/zenodo.21803258 (Lippert et al., 2026b).

Appendix A: Jacobian evaluation

Fig. A1-A3 show the agreement between emulator and full-physics Jacobians as discussed in Section 3.1.3 in the main text.

Appendix B: Additional retrieval results

Figure B1 shows scatter plots of retrieved albedo and aerosol parameters. With all forward model variants, albedo parameters are retrieved with high accuracy (RMSE ≤ 0.01). As expected, aerosol parameters are more difficult to retrieve from spectrometer measurements alone, but all emulators are able to provide meaningful information on both optical depth (RMSE

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

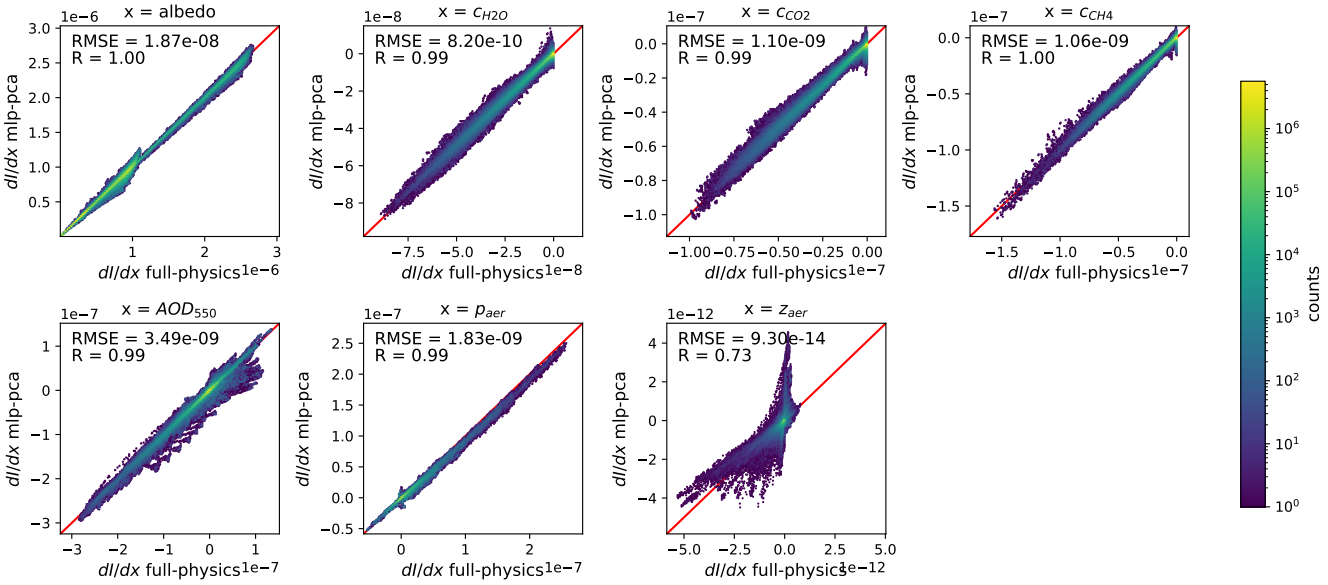

Figure A1. Jacobian evaluation for the base emulator. Each subplot shows partial derivatives dI/dx of predicted radiances w.r.t. input parameter x against the corresponding partial derivatives of the full-physics model. All instrument channels across the NIR, SWIR1, and SWIR3 bands are plotted together. The red 1:1 line represents perfect alignment with the full-physics model.

≤ 0.07) and particle size (RMSE ≤ 0.61). Similar to XCH_4 and XCO_2 retrievals, the base emulator results in the largest aerosol retrieval errors, while the two hybrid emulators achieve accuracies close to the full-physics reference.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

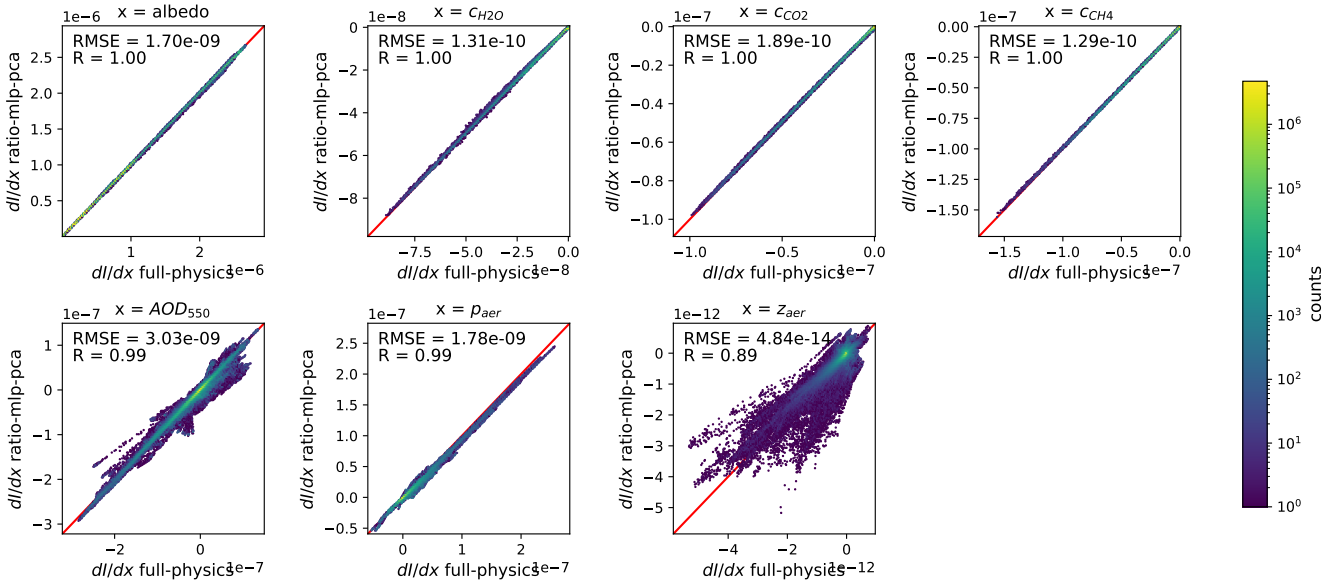

Figure A2. Jacobian evaluation for the hybrid MLP-PCA emulator. Each subplot shows partial derivatives dI/dx of predicted radiances w.r.t. input parameter x against the corresponding partial derivatives of the full-physics model. All instrument channels across the NIR, SWIR1, and SWIR3 bands are plotted together. The red 1:1 line represents perfect alignment with the full-physics model.

Author contributions. FL, AGB, and JL conceived the study and developed the methodology. FL implemented, trained, and analysed the neural network models, and integrated them into the RemoTAP retrieval code with help from SL and OH. FL wrote the original draft. All authors contributed to the discussion of the results and to the review and editing of the final manuscript.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Atmospheric Measurement Techniques.

Acknowledgements. This research was performed within the framework of the ESA OPERA-S5 project (grant no. 4000143572/24/I-KE). Neural network training was performed using the Dutch national e-infrastructure with the support of the SURF Cooperative (grant no. EINF-11809). Large Language models (OpenAI's GPT-5.5 and Anthropic's Claude Sonnet 5) were used for English language editing, proofreading, and improving the overall clarity of the manuscript.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

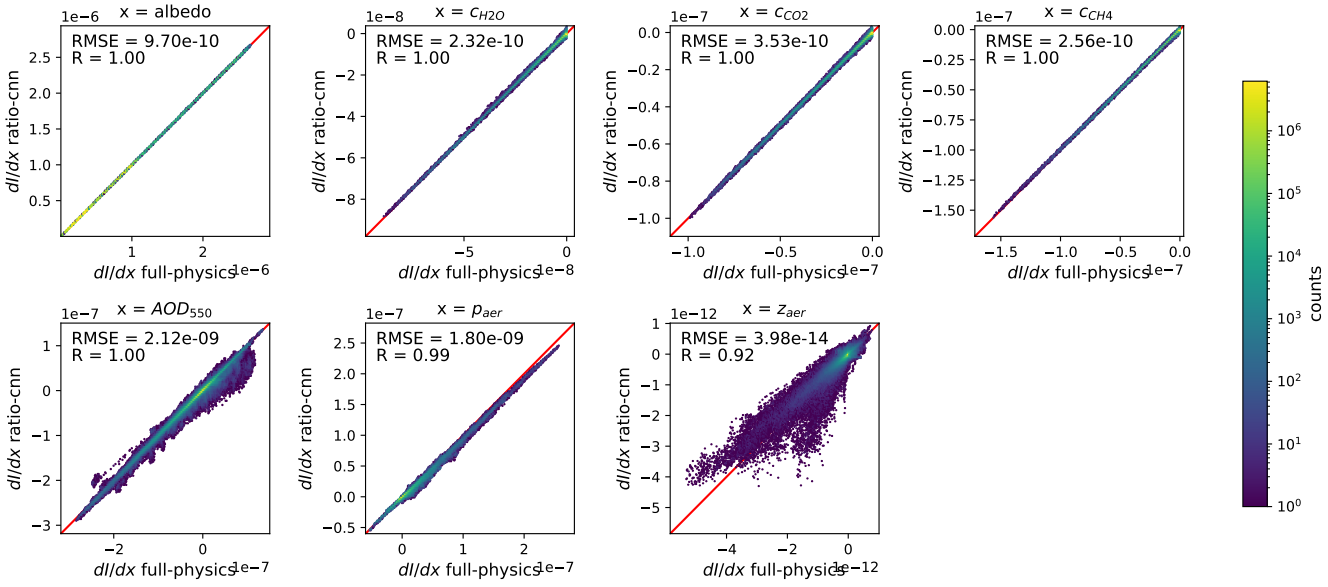

Figure A3. Jacobian evaluation for the hybrid CNN emulator. Each subplot shows partial derivatives dI/dx of predicted radiances w.r.t. input parameter x against the corresponding partial derivatives of the full-physics model. All instrument channels across the NIR, SWIR1, and SWIR3 bands are plotted together. The red 1:1 line represents perfect alignment with the full-physics model.

References

Anderson, G. P., Clough, S. A., Kneizys, F. X., Chetwynd, J. H., and Shettle, E. P.: AFGL atmospheric constituent profiles (0.120km), 1986.

Atkinson, J., Elafrou, A., Kasoar, E., Wallwork, J. G., Meltzer, T., Clifford, S., Orchard, D., and Edsall, C.: FTorch: a library for coupling PyTorch models to Fortran, Journal of Open Source Software, 10, 7602, 2025.

Borsdorff, T., Hasekamp, O., Wassmann, A., and Landgraf, J.: Insights into Tikhonov regularization: application to trace gas column retrieval and the efficient calculation of total column averaging kernels, Atmospheric Measurement Techniques, 7, 523–535, 2014.

Brence, J., Tanevski, J., Adams, J., Malina, E., and Džeroski, S.: Surrogate models of radiative transfer codes for atmospheric trace gas retrievals from satellite observations, Mach Learn, 112, 1337–1363, https://doi.org/10.1007/s10994-022-06155-2, 2023.

Brodrick, P. G., Thompson, D. R., Fahlen, J. E., Eastwood, M. L., Sarture, C. M., Lundeen, S. R., Olson-Duvall, W., Carmon, N., and Green, R. O.: Generalized radiative transfer emulation for imaging spectroscopy reflectance retrievals, Remote Sensing of Environment, 261, 112 476, https://doi.org/10.1016/j.rse.2021.112476, 2021.

Bue, B. D., Thompson, D. R., Deshpande, S., Eastwood, M., Green, R. O., Natraj, V., Mullen, T., and Parente, M.: Neural network radiative transfer for imaging spectroscopy, Atmospheric Measurement Techniques, 12, 2567–2578, https://doi.org/10.5194/amt-12-2567-2019, 2019.

Butz, A., Hasekamp, O. P., Frankenberg, C., and Aben, I.: Retrievals of atmospheric CO 2 from simulated space-borne measurements of backscattered near-infrared sunlight: accounting for aerosol effects, Applied optics, 48, 3322–3336, 2009.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

Butz, A., Hasekamp, O., Frankenberg, C., Vidot, J., and Aben, I.: CH4 retrievals from space-based solar backscatter measurements: Performance evaluation against simulated aerosol and cirrus loaded scenes, Journal of Geophysical Research: Atmospheres, 115, 2010.

Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze, A., et al.: Toward accurate CO2 and CH4 observations from GOSAT, Geophysical Research Letters, 38, 2011.

Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO2 sources and sinks: Theoretical study in a variational data assimilation framework, Journal of Geophysical Research: Atmospheres, 112, 2007.

Clough, S. A., Shephard, M. W., Mlawer, E. J., Delamere, J. S., Iacono, M. J., Cady-Pereira, K., Boukabara, S., and Brown, P. D.: Atmospheric radiative transfer modeling: A summary of the AER codes, Journal of Quantitative Spectroscopy and Radiative Transfer, 91, 233–244, 2005.

David, L., Bréon, F.-M., and Chevallier, F.: XCO 2 estimates from the OCO-2 measurements using a neural network approach, Atmospheric Measurement Techniques, 14, 117–132, 2021.

Efremenko, D., Doicu, A., Loyola, D., and Trautmann, T.: Optical property dimensionality reduction techniques for accelerated radiative transfer performance: Application to remote sensing total ozone retrievals, Journal of Quantitative Spectroscopy and Radiative Transfer, 133, 128–135, 2014.

Eldering, A., Wennberg, P., Crisp, D., Schimel, D., Gunson, M., Chatterjee, A., Liu, J., Schwandner, F., Sun, Y., O'dell, C., et al.: The Orbiting Carbon Observatory-2 early science investigations of regional carbon dioxide fluxes, Science, 358, eaam5745, 2017.

Elfwing, S., Uchibe, E., and Doya, K.: Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, Neural networks, 107, 3–11, 2018.

Fan, C., Fu, G., Di Noia, A., Smit, M., H.H. Rietjens, J., A. Ferrare, R., Burton, S., Li, Z., and P. Hasekamp, O.: Use of A Neural Network-Based Ocean Body Radiative Transfer Model for Aerosol Retrievals from Multi-Angle Polarimetric Measurements, Remote Sensing, 11, 2877, https://doi.org/10.3390/rs11232877, 2019.

Frankenberg, C., Platt, U., and Wagner, T.: Iterative maximum a posteriori (IMAP)-DOAS for retrieval of strongly absorbing trace gases: Model studies for CH 4 and CO 2 retrieval from near infrared spectra of SCIAMACHY onboard ENVISAT, Atmospheric Chemistry and Physics, 5, 9–22, 2005.

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., et al.: Global carbon budget 2023, Earth System Science Data, 15, 5301–5369, 2023.

Fu, Q. and Liou, K.: On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres, Journal of Atmospheric Sciences, 49, 2139–2156, 1992.

Gao, M., Franz, B. A., Knobelspiesse, K., Zhai, P.-W., Martins, V., Burton, S., Cairns, B., Ferrare, R., Gales, J., Hasekamp, O., Hu, Y., Ibrahim, A., McBride, B., Puthukkudy, A., Werdell, P. J., and Xu, X.: Efficient multi-angle polarimetric inversion of aerosols and ocean color powered by a deep neural network forward model, Atmospheric Measurement Techniques, 14, 4083–4110, https://doi.org/10.5194/amt-14-4083-2021, 2021.

Gordon, I. E., Rothman, L. S., Hargreaves, e. R., Hashemi, R., Karlovets, E. V., Skinner, F., Conway, E. K., Hill, C., Kochanov, R. V., Tan, Y., et al.: The HITRAN2020 molecular spectroscopic database, Journal of quantitative spectroscopy and radiative transfer, 277, 107 949, 2022.

Guerlet, S., Butz, A., Schepers, D., Basu, S., Hasekamp, O., Kuze, A., Yokota, T., Blavier, J.-F., Deutscher, N., Griffith, D. T., et al.: Impact of aerosol and thin cirrus on retrieving and validating XCO2 from GOSAT shortwave infrared measurements, Journal of Geophysical Research: Atmospheres, 118, 4887–4905, 2013.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

Gómez-Dans, J. L., Lewis, P. E., and Disney, M.: Efficient Emulation of Radiative Transfer Codes Using Gaussian Processes and Application to Land Surface Parameter Inferences, Remote Sensing, 8, 119, https://doi.org/10.3390/rs8020119, number: 2, 2016.

Hansen, J. E. and Travis, L. D.: Light scattering in planetary atmospheres, Space science reviews, 16, 527–610, 1974.

Hasekamp, O. P. and Butz, A.: Efficient calculation of intensity and polarization spectra in vertically inhomogeneous scattering and absorbing atmospheres, Journal of Geophysical Research: Atmospheres, 113, 2008.

Hasekamp, O. P. and Landgraf, J.: Linearization of vector radiative transfer with respect to aerosol properties and its use in satellite remote sensing, Journal of Geophysical Research: Atmospheres, 110, 2005.

Hasekamp, O. P., Litvinov, P., and Butz, A.: Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements, Journal of Geophysical Research: Atmospheres, 116, 2011.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

Heiskanen, J., Brümmer, C., Buchmann, N., Calfapietra, C., Chen, H., Gielen, B., Gkritzalis, T., Hammer, S., Hartman, S., Herbst, M., et al.: The integrated carbon observation system in Europe, Bulletin of the American Meteorological Society, 103, E855–E872, 2022.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural networks, 2, 359–366, 1989.

Hu, H., Hasekamp, O., Butz, A., Galli, A., Landgraf, J., Aan de Brugh, J., Borsdorff, T., Scheepmaker, R., and Aben, I.: The operational methane retrieval algorithm for TROPOMI, Atmospheric Measurement Techniques, 9, 5423–5440, 2016.

Hu, H., Landgraf, J., Detmers, R., Borsdorff, T., Aan de Brugh, J., Aben, I., Butz, A., and Hasekamp, O.: Toward global mapping of methane with TROPOMI: First results and intersatellite comparison to GOSAT, Geophysical Research Letters, 45, 3682–3689, 2018.

Ingmann, P., Veihelmann, B., Langen, J., Lamarre, D., Stark, H., and Courrèges-Lacoste, G. B.: Requirements for the GMES Atmosphere Service and ESA's implementation concept: Sentinels-4/-5 and -5p, Remote Sensing of Environment, 120, 58–69, https://doi.org/https://doi.org/10.1016/j.rse.2012.01.023, the Sentinel Missions - New Opportunities for Science, 2012.

IPCC: Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2021.

Irizar, J., Melf, M., Bartsch, P., Koehler, J., Weiss, S., Greinacher, R., Erdmann, M., Kirschner, V., Albinana, A. P., and Martin, D.: Sentinel-5/UVNS, in: International Conference on Space Optics — ICSO 2018, edited by Sodnik, Z., Karafolas, N., and Cugny, B., vol. 11180, p. 1118004, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.2535923, 2019.

Kingma, D. P. and Ba, J. L.: Adam: A method for stochastic optimization, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pp. 1–15, https://arxiv.org/abs/1412.6980, 2015.

Lacis, A. A. and Oinas, V.: A description of the correlated k distribution method for modeling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres, Journal of Geophysical Research: Atmospheres, 96, 9027–9063, 1991.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

Lamminpää, O., Susiluoto, J., Hobbs, J., McDuffie, J., Braverman, A., and Owhadi, H.: Forward model emulator for atmospheric radiative transfer using Gaussian processes and cross validation, Atmospheric Measurement Techniques, 18, 673–694, https://doi.org/10.5194/amt-18-673-2025, 2025.

Landgraf, J., Hasekamp, O., and Trautmann, T.: Linearization of radiative transfer with respect to surface properties, Journal of Quantitative Spectroscopy and Radiative Transfer, 72, 327–339, 2002.

Landgraf, J., Butz, A., Hasekamp, O., Hu, H., and aan de Brugh, J.: Sentinel 5 L2 Prototype Processors, Algorithm Theoretical Baseline Document: Methane Retrieval, ATBD SRON-ESA-S5L2PP-ATBD-001, Space Research Organisation Netherlands (SRON), https://user.eumetsat.int/s3/ope-eup-strapi-media/EPS_SG_Sentinel_5_Level_2_Methane_ATBD_6c6020b98b.pdf, version 5.3, 2025.

Langen, J., Meijer, Y., Brinksma, E., Veihelmann, B., and Ingmann, P.: GMES Sentinels 4 and 5 mission requirements document, Eop-sma/1507/jl-dr, ESA ESRIN, 2011.

Lenoble, J. et al.: Radiative transfer in scattering and absorbing atmospheres: standard computational procedures, vol. 300, A. Deepak Hampton, Va., 1985.

Lippert, F.: FionaLippert/Surrogate_RTM: v1.1, https://doi.org/10.5281/zenodo.21772755, 2026.

Lippert, F., Barr, A. G., Herreras-Giralda, M., Momoi, M., Rejano, F., Lu, S., Hasekamp, O., Dubovik, O., Malina, E., and Landgraf, J.: Data from: Accelerating greenhouse gas retrievals with neural network-based forward models, https://doi.org/10.5281/zenodo.21792065, 2026a.

Lippert, F., Barr, A. G., Herreras-Giralda, M., Momoi, M., Rejano, F., Lu, S., Hasekamp, O., Dubovik, O., Malina, E., and Landgraf, J.: Synthetic Sentinel-5/UVNS measurements, https://doi.org/10.5281/zenodo.21803258, 2026b.

Liu, X., Smith, W. L., Zhou, D. K., and Larar, A.: Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept, Applied Optics, 45, 201–209, 2006.

Lorente, A., Borsdorff, T., Butz, A., Hasekamp, O., aan de Brugh, J., Schneider, A., Wu, L., Hase, F., Kivi, R., Wunch, D., et al.: Methane retrieved from TROPOMI: improvement of the data product and validation of the first 2 years of measurements, Atmospheric Measurement Techniques, 14, 665–684, 2021.

Luo, D., O'Leary-Roseberry, T., Chen, P., and Ghattas, O.: Efficient PDE-constrained optimization under high-dimensional uncertainty using derivative-informed neural operators, SIAM Journal on Scientific Computing, 47, C899–C931, 2025.

Mauceri, S., O'Dell, C. W., McGarragh, G., and Natraj, V.: Radiative Transfer Speed-Up Combining Optimal Spectral Sampling With a Machine Learning Approach, Front. Remote Sens., 3, 932 548, https://doi.org/10.3389/frsen.2022.932548, 2022.

Meirink, J., Eskes, H., and Goede, A.: Sensitivity analysis of methane emissions derived from SCIAMACHY observations through inverse modelling, Atmospheric Chemistry and Physics, 6, 1275–1292, 2006.

Moncet, J.-L., Uymin, G., Lipton, A. E., and Snell, H. E.: Infrared radiance modeling by optimal spectral sampling, Journal of the Atmospheric Sciences, 65, 3917–3934, 2008.

Nanda, S., de Graaf, M., Veefkind, J. P., ter Linden, M., Sneep, M., de Haan, J., and Levelt, P. F.: A neural network radiative transfer model approach applied to the Tropospheric Monitoring Instrument aerosol height algorithm, Atmospheric Measurement Techniques, 12, 6619–6634, https://doi.org/10.5194/amt-12-6619-2019, 2019.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., et al.: Automatic differentiation in PyTorch, Advances in Neural Information Processing Systems, 31, 2017.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library, Advances in Neural Information Processing Systems, 33, 2019.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

Phillips, D. L.: A technique for the numerical solution of certain integral equations of the first kind, Journal of the ACM (JACM), 9, 84–97, 1962.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, vol. 2, World scientific, 2000.

Schepers, D., Aan de Brugh, J., Hahne, P., Butz, A., Hasekamp, O., and Landgraf, J.: LINTRAN v2. 0: A linearised vector radiative transfer model for efficient simulation of satellite-born nadir-viewing reflection measurements of cloudy atmospheres, Journal of Quantitative Spectroscopy and Radiative Transfer, 149, 347–359, 2014.

Schuit, B. J., Maasakkers, J. D., Bijl, P., Mahapatra, G., Van den Berg, A.-W., Pandey, S., Lorente, A., Borsdorff, T., Houweling, S., Varon, D. J., et al.: Automated detection and monitoring of methane super-emitters using satellite data, Atmospheric Chemistry and Physics Discussions, 2023, 1–47, 2023.

Sierk, B., Bezy, J.-L., Caron, J., Meynard, R., Veihelmann, B., and Ingmann, P.: The GMES Sentinel-5 mission for operational atmospheric monitoring: status and developments, in: International Conference on Space Optics — ICSO 2012, edited by Cugny, B., Armandillo, E., and Karafolas, N., vol. 10564, p. 1056434, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.2309276, 2018.

Tikhonov, A. N.: On the solution of incorrectly stated problems and a method of regularization, in: Dokl. Acad. Nauk. SSSR, vol. 151, pp. 501–504, 1963.

Verrelst, J., Rivera Caicedo, J. P., Muñoz-Marí, J., Camps-Valls, G., and Moreno, J.: SCOPE-Based Emulators for Fast Generation of Synthetic Canopy Reflectance and Sun-Induced Fluorescence Spectra, Remote Sensing, 9, 927, https://doi.org/10.3390/rs9090927, number: 9, 2017.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W., Sherlock, V., and Wennberg, P. O.: The total carbon column observing network, Philosophical transactions of the Royal Society A: Mathematical, physical and engineering sciences, 369, 2087–2112, 2011.

https://doi.org/10.5194/egusphere-2026-4724
Preprint. Discussion started: 12 August 2026

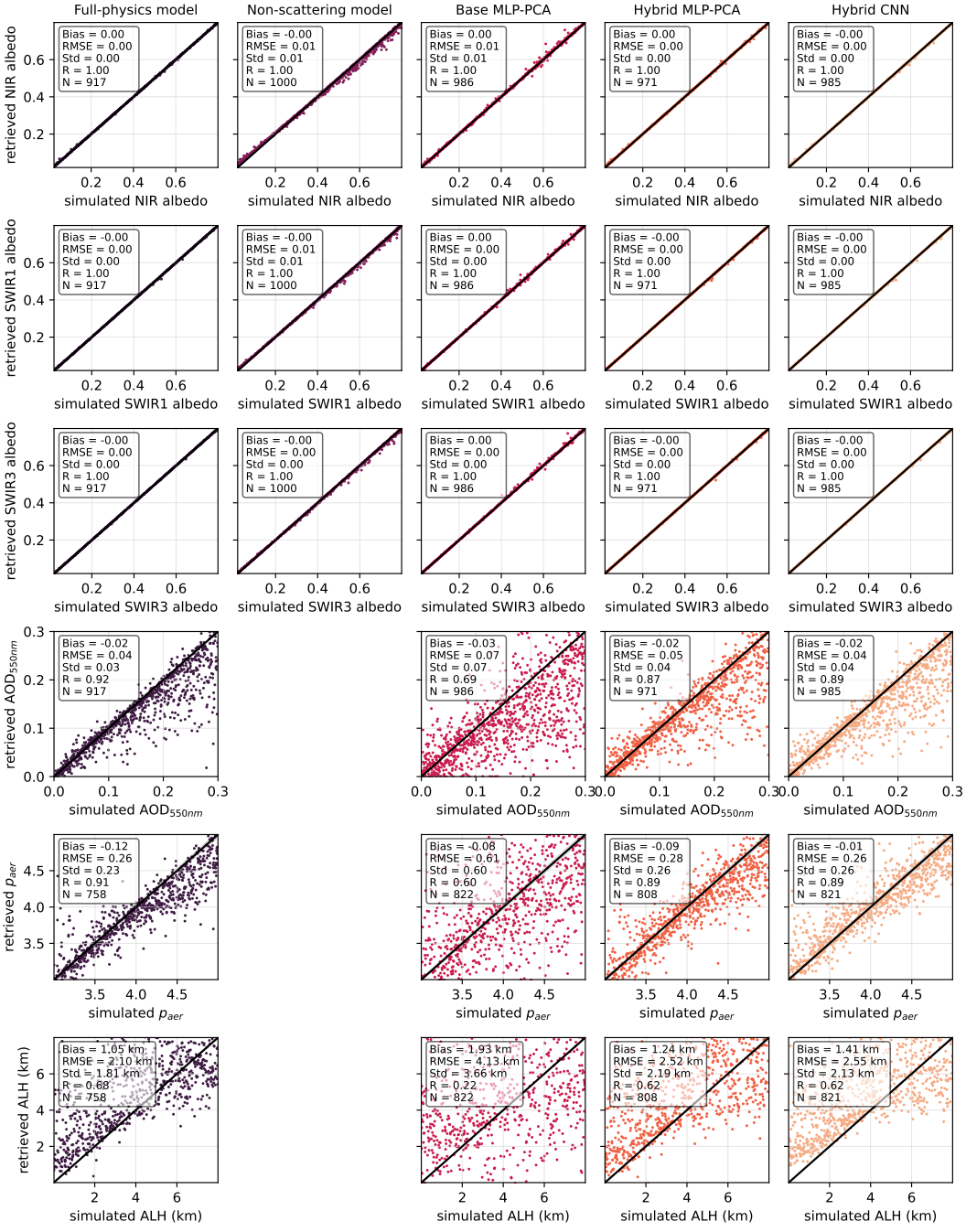

Figure B1. Retrieved vs. simulated albedo and aerosol parameters. For the aerosol size parameter p_{aer} and aerosol layer hight (ALH), only scenes with AOD_{550nm} larger than 0.05 are taken into account. For non-scattering forward model, aerosol parameters are not shown as they are not included in the state vector.